

Chapter 2

Volatility Dynamics for a Single Underlying: Foundations

Abstract In this first and fundamental chapter we lay out the core principles of the Asymptotic Chaos Expansion (ACE) methodology. We investigate the relationship between *stochastic instantaneous volatility* (SInsV) and *stochastic implied volatility* (SImpV) models, in the simple case of a single underlying, and when the endogenous driver is scalar. We discuss both the inverse (or recovery) and the direct problem, initially limiting the asymptotic expansion to its lowest order, which we call the *first layer*. We illustrate these asymptotic results first with the local volatility (LV) class, and then with a comprehensive extension to stochastic volatility (SV) dynamics.

In Sect. 2.1, we define the market environment: the underlying, the numeraire and the liquid European options. We define and justify the re-parametrisation of the option price surface *via* a *sliding* implied volatility map. We can then introduce both stochastic volatility models (SInsV and SImpV) as well as some sufficient regularity assumptions. Finally we state our objectives, which we split into a *direct* and an *inverse* problem.

In Sect. 2.2, we establish the fundamental result of this chapter, and of the ACE methodology. This is the *Zero Drift Condition* (ZDC), a PDE constraining the shape and dynamics of the stochastic *implied* volatility model in the whole strike/expiry domain, in order to respect the no-arbitrage assumption (NAA). We then specialise that result to the immediate or zero-expiry sub-domain, which leads us to a pair of *Immediate ZDCs*. Finally we specialise again these results to the *Immediate At-The-Money* (IATM) point, which is our most limited but fertile asymptotic, and quote the IATM Identity linking the implied and instantaneous volatilities.

In Sect. 2.3, we solve part of the *inverse* problem, which is to recover the *instantaneous* model from the *implied* one. First we establish arbitrage constraints between the coefficients of the SImpV model at the IATM point, which emphasises the structural over-specification of that class. Then we show that at a given level of precision (the *first layer*, which involves a group of low-order IATM differentials of the smile) the implied model injects itself into the instantaneous class.

In Sect. 2.4, we tackle the more popular *direct* problem, which consists in generating the smile, and more generally the *implied* model associated to a given *instantaneous* class. For the first layer, we establish the opposite connection from before, which confirms a full correspondence between the two classes (at that level of

precision). We comment, illustrate and contrast these results against the available literature, within the simple class of local volatility models, thereby exposing some shortcomings in a popular heuristic.

In Sect. 2.5 we turn to some practical applications of these results, which we classify as either pure asymptotic, whole-smile or sensitivity-oriented. In terms of pure asymptotics, we define a stochastic instantaneous volatility model class, covering most popular SV models, for which we provide the first layer differentials. In particular we analyse the respective merits and properties of the Lognormal displaced diffusion (LDD) and CEV specifications as skew functions. We also use the pure asymptotic results to re-parametrise such instantaneous volatility models into more *intuitive* versions, based on the smile that they generate. Then we briefly discuss the caveats involved in extrapolating the whole smile in the naive way, *via* Taylor expansions.

Eventually, in Sect. 2.6 we conclude this chapter and open onto the more specialised subjects covered thereafter. We also provide a diagram gathering the main proofs of this chapter, and hence capturing the main body of the ACE approach, as a basis and a comparison tool for the more advanced results coming thereafter. Note that a general roadmap for this chapter is provided by Fig. 2.4 [p. 113].

2.1 Framework and Objectives

2.1.1 Market and Underlyings

We consider a market equipped with the usual filtered objective probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$. Unless explicitly specified otherwise, all processes mentioned thereafter will be assumed to be continuous and $\mathcal{F}_t$ -adapted.

Although we impose the No-Arbitrage Assumption (NAA), we do not, however, demand market completeness. This choice is obviously motivated by the stochastic volatility specification, and therefore in the sequel the term “risk-neutral measure” should be understood as “chosen risk-neutral measure” with respect to the *volatility risk premium*. Some considerable literature has been devoted to the economic significance and modelling of this risk premium. We have chosen not to dwell on this interesting subject, since it is of less relevance in our “completed” framework (which includes vanilla options), and also because it presents less interest in the prospect of pure hedging (using dynamically those options). Those readers interested in the subjects of market equilibrium, and market price of volatility risk, can refer to [1] or [2], for instance.

As for the theoretical market, we start by considering a single, scalar asset, with continuous price process $\widehat{S}_t$. We also select a numeraire asset N_t , so that under the risk-neutral measure $\mathbb{Q}$ and using Lognormal conventions, their dynamics come as

$$d\widehat{S}_t/\widehat{S}_t = r_t dt + \sigma_t dW_t^{\mathbb{Q}} \quad \text{and} \quad dN_t/N_t = r_t dt + \lambda_t dW_t^{\mathbb{Q}}$$

with $W_t^{\mathbb{Q}}$ a scalar $\mathbb{Q}$ -Wiener process, while both σ_t and λ_t are undefined but continuous scalar volatility processes. We then define the *deflated or rebased asset* S_t and its martingale measure $\mathbb{Q}^N$ by writing

$$S_t = \widehat{S}_t / N_t \quad \text{and} \quad dW_t^{\mathbb{Q}^N} = dW_t^{\mathbb{Q}} - \lambda_t dt$$

and specifying that $W_t^{\mathbb{Q}^N}$ be a $\mathbb{Q}^N$ -Wiener process. But now the process S_t is not *a priori* a tradeable asset any more. For all intents and purposes, it should be seen as an index, a reference defining the payoffs of our soon-to-come vanilla options. For that reason, it will be called an *underlying*, with dynamics coming driftless as

$$\frac{dS_t}{S_t} = \sigma_t dW_t^{\mathbb{Q}^N}. \quad (2.1)$$

Nevertheless, as a matter of convention S_t will often be called the “spot” process in the sequel. Also, regarding notation, we will forgo the $\mathbb{Q}^N$ identifier for the relevant Wiener process and simply denote it by W_t .

2.1.2 Vanilla Options Market and Sliding Implied Volatility

2.1.2.1 Definitions and Notations

On top of the underlying S_t and of the numeraire N_t , we now assume a *market continuum* of prices $C(t, S_t, T, K)$ for Call options written on S_t . Their payoff is either defined or equivalent (using NAA arguments) to the following cashflow, transferred at time T :

$$C(T, S_T, T, K) = N_T (S_T - K)^+. \quad (2.2)$$

The continuum is assumed both in maturity (until a finite horizon $T_{\max}$) and in strike (for all $K \in]-\infty, \infty[$). In fact we could consider Puts or Straddles instead of Calls: a smooth continuum in strike is indeed equivalent to assuming that the full marginal distribution is given.

Providing it is *valid*,¹ this surface of option prices is associated to an implied volatility mapping $\Sigma(t, S_t, K, T)$ via the classical Lognormal re-parametrisation:

$$C(t, S_t, T, K) = N_t C^{BS} \left(S_t, K, \Sigma(t, S_t, K, T) \cdot \sqrt{T - t} \right) \quad (2.3)$$

¹ In other words, that the implied marginal densities satisfy the usual criteria, see Sect. 4.1.

where $C^{BS}(x, k, v)$ is the *time-normalised* Black functional (see [3]), which we now define. Denoting by

$$y(x, k) \triangleq \ln(k/x)$$

the *log-strike* relative to the spot, *a.k.a.* “log-moneyness”, we set

$$C^{BS} : \mathbb{R}^{+2} \times \mathbb{R}^{+*} \rightarrow \mathbb{R}^+ \quad C^{BS}(x, k, v) \triangleq x\mathcal{N}(d_1) - k\mathcal{N}(d_2) \quad (2.4)$$

$$\text{with} \quad d_{1/2}(x, k, v) = \frac{-y}{v} \pm \frac{1}{2}v \quad \text{and} \quad \mathcal{N}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}s^2} ds.$$

Note that, since both the Calls and the numeraire have been chosen as traded assets, their ratio as per (2.3) will naturally be martingale under $\mathbb{Q}^N$: it is the numeraire N_t that ensures the necessary link between the underlying and the payoff ((2.1) vs (2.2)). The whole construction (asset, numeraire, payoff) will appear arch-classical to any reader familiar with the interest rates environment, and might even look restrictive. In practice, however, it covers most existing vanilla products and market conventions.

Besides, it is possible to extend this simple framework to less classical configurations. For instance, we could theoretically deal with different drifts between the “asset”, the numeraire and the Calls. We could also consider different pricing functionals from the Black formula, and/or look at other payoff definitions than Call options.

Most of these possibilities will be discussed in Chap. 3, which deals with extensions of the basic framework. Some of these configurations, in particular drift misalignment, will be used out of necessity in Chap. 5, dealing with the term structure framework. But all in all, the basic setup that we consider here is a good starting point, simply because, by killing the drift, it will enable us to derive shortly a clean, simple Zero-Drift Condition (2.18), which is the foundation of our results.

As will be made clearer in Sect. 4.1, the validity of the *price* mapping itself is reasonably simple to establish. The static aspect for instance can often be checked visually. However, doing so through the Implied Volatility re-parametrisation is quite technical and can prove counterintuitive. Therefore, for the moment, we will put that issue aside and simply assume that the IV surface is statically and dynamically valid.

We now associate to these “absolute” quantities $C(\cdot)$ and $\Sigma(\cdot)$ their “sliding” counterparts, respectively $\tilde{C}(\cdot)$ and $\tilde{\Sigma}(\cdot)$. Let us recall that at any given time a sliding quantity can be made to match an absolute one, but that their *dynamics* will thereafter diverge and will therefore provide different insight. These new maps are parametrised w.r.t. a couple of new quantities:

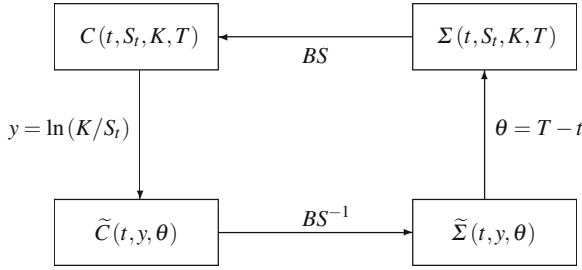


Fig. 2.1 Option price and implied volatility: absolute vs sliding

The <i>time-to-maturity</i> ²	$\theta \triangleq (T - t).$
The <i>log-moneyness</i>	$y \triangleq \ln(K/S_t).$

Beware that the log-moneyness is defined here as the opposite of $\ln(S_t/K)$, which itself tends to be found in many related papers (since it represents a term of the Black Call pricing formula). Formally we write (see Fig. 2.1)

$$C(t, S_t, K, T) \triangleq \tilde{C}(t, y, \theta) \quad \text{and} \quad \Sigma(t, S_t, K, T) \triangleq \tilde{\Sigma}(t, y, \theta).$$

More generally, we will use the superscript $\tilde{\cdot}$ to identify all sliding quantities, in strike and/or in maturity. However, it must be understood as simultaneously affecting both coordinates, if these are present among the arguments.

Of particular interest in the sliding representation are two regions of the map:

Immediate	will refer to	$\theta \equiv 0.$
At-The-Money (ATM)	corresponds naturally to	$y \equiv 0.$

In our asymptotic context, the intersection of both domains is pivotal, hence we denote by **IATM** (Immediate ATM) the point $(\mathbf{y} = \mathbf{0}, \theta = \mathbf{0})$.

Since a large part of this study will be spent differentiating absolute and sliding functionals with respect to their arguments, it makes sense to gather in a single place all transition formulae between the two configurations: this is the object of Appendix B [p. 431].

² “TTM” in shorthand.

2.1.2.2 Motivation and Properties of the Black IV Representation

Since the Black formula assumes Lognormal dynamics for the underlying asset, re-parameterising with the normalised BS implied volatility seems appropriate when S_t is not only martingale (under the measure associated to N_t) but also exhibits “close to Lognormal” dynamics. In some practical instances, other simple dynamics such as the *Normal* framework³ can prove efficient, as will be discussed in Sect. 3.3.3. But, in most markets, the support of the asset marginal distribution is constrained (or assumed) to be asymmetric, typically bounded on the left. Therefore the (displaced) Black-Scholes implied volatility has proven to be a robust⁴ candidate for the re-parametrisation of the price map.

In a more general manner, it is in fact the “implied parameter” approach, which consists in considering prices through a simple “baseline” model, which allows us to compare “raw” prices for different strikes and/or maturities. The Normal and Lognormal dynamics are merely instances of that approach, albeit very common and important ones (more on this in Sect. 3.3.3).

Another advantage of the IV map over the price map is its *regularisation effect*, which is ironically a consequence of its more limited domain of definition.

Indeed, the Black-Scholes formula (2.4) is only specified in the domain $\theta > 0$, since at $T = t$ the option price naturally equals the intrinsic payoff $[S_T - K]^+$. The latter, however, does not provide C^1 regularity at-the-money. This is an issue since our method happens to be of an *asymptotic* nature. It uses expansions intensively and therefore requires/provides differentials of some transform of the price, taken precisely at that same IATM point $(t, y = 0, \theta = 0)$. Alternatively, if we assume that the implied volatility is well behaved for short maturities, typically if it admits a finite limit in $T = t$ along with a sufficient number of its differentials, we can extend the IV map by continuity.⁵ The Black-Scholes formula itself becomes valid in the full domain and allows us to fall back effortlessly onto the intrinsic value. This re-parametrisation effectively *contains* the irregularity of the price functional to the Black formula itself, allowing the new functional (the implied volatility) to be infinitely smooth, if required.

In other words, the price always exhibit a singularity (a “kink”) at the IATM point, while the implied volatility *can* be infinitely smooth. It then becomes clear that an additional and major attraction of re-parameterising with the Implied Volatility is that it enables, at low cost, the local regularity that our methodology requires.

In the same vein, the expansion method that we use is necessarily less precise for strikes far from the money. Therefore, since the Vega dies out in these regions, using volatility (as opposed to price) expansions artificially limits the resulting *pricing* error, which is most important trading-wise. In other words, in terms of magnitude

³ And therefore the Bachelier formula.

⁴ For other definitions and sources of robustness for Black-Scholes, see [4], for instance.

⁵ It is for this reason that, in the sequel, any value of the implied volatility taken with $\theta = 0$, typically in $(t, y, 0)$ or $(t, 0, 0)$, must be viewed as a notation abuse, in fact a limit.

the IV is usually a more uniform, precise albeit dangerous (*c.f.* validity issues) representation than the price itself.

To complement this point, it is also interesting to note that the implied volatility is in practice strongly linked to the Delta, hence to the hedge, and especially so in the FX world.

As a final word of caution, we stress that the argumentation above is valid for vanilla Call and Put options, but might not be so for other products, such as binaries: there will be more on this point in Sect. 3.3.3.

2.1.2.3 Motivation and Properties of the Sliding Representation

We first put this technique into perspective and comment on its relevant mathematical properties. We then discuss the financial attractiveness of this simultaneous “time and strike” slide.

The general concept and the use of relative variables are certainly not new. In the rates environment for instance, it is common practice to denote a Libor rate either with fixed maturity or with fixed accrual⁶: each notation has its specific pros and cons (see [5] or [6], among others). In an option framework, sliding strikes are also frequently used in order to account for “stickiness”: certain smiles are “strike sticky” while most are “Delta sticky”.

Besides, we emphasise that the nature of the benefit brought by this sliding convention, *in our specific framework*, is more style than substance. Indeed, it does not lead to fundamental or technical results which an absolute setting could not reach. This is a positive feature, since our choice of a strike representation (log-moneyness) is partly subjective and certainly no panacea. It is therefore comforting that our results can effectively be transferred to another convention: the practical aspects of this move are discussed in Sect. 2.1.3.2.

It remains that in principle there are many such ways to define the slide, especially in strike. An obvious candidate is proportional moneyness (K/S_t), but any other adequate function of K and S_t can be considered: such adequacy obviously requires a bijectivity in K and also a sufficient regularity, especially at the money. In [7] one can find a general definition for the strike slide, called simply *moneyness*. But it is stressed therein that the choice should be made on an *ad hoc* basis, an assertion that we support. Indeed, for a given market, a good parametrisation should provide a smile dynamically as stable and stationary as possible. The overall principle consists in conditioning the smile w.r.t. our only observable state variables, i.e. t and S_t . For a complementary discussion on this subject, refer to Sect. 3.3.1.1 [p. 148].

We believe, however, that our specific choice of a sliding convention is justified, for reasons that we expose now. First of all, and on a mathematical level, we elected to use Lognormal dynamics to define an implied parameter: this is in no way mandatory and is simply the most common market practice. However, it leads to Black’s formula, which itself clearly makes of the log-moneyness y the *natural* variable to consider.

⁶ $L(t, T, U)$ vs $L^\delta(t, T) \triangleq L(t, T, T + \delta)$.

Also, the results that we present herein are structurally complex, hence any approach that clarifies the interpretation and the role of the various terms is *a priori* welcome. In particular, we find that the sliding representation usually allows a better understanding of stationarity and time-homogeneity issues, especially in the second Part devoted to term structure models.

Furthermore, and as will be covered in the various application sections, the practical efficiency of our methodology depends as much on the pure asymptotic results as on the chosen representation of the variables. This is generally true for most extrapolation methods, but also for numerical reasons as well as for the analysis of model behaviour. For the latter in particular, when the model and/or smile specification themselves are (pseudo-) sliding or time-shift homogeneous, we find that better efficiency is attained by using the sliding versions of our results.

In more formal terms, the main attraction towards sliding versions of the price and implied volatility mappings lies in the dynamic and stochastic properties brought by this change of coordinates. Indeed, reducing the number of arguments from an absolute representation (four arguments: t, S_t, T and K) to a sliding one (three arguments only: t, y and θ) effectively “transfers” the underlying S_t , and therefore the driver W_t into the functional (here $\tilde{S}$ or $\tilde{C}$). Let us quickly illustrate this point with two simple examples.

Example 2.1 (Markovian dimension of the IV for a pure local volatility (LV) model) First let us assume that the smile is generated by a pure LV model, as defined by

$$\frac{dS_t}{S_t} = f(t, S_t) dW_t \quad \text{with} \quad f(s, x) \geq 0 \forall (s, x) \in \mathbb{R}^{+2}. \quad (2.5)$$

Then the Markovian state variables are simply t and S_t , so that both the absolute price surface $C(t, S_t; K, T)$ and the implied volatility surface $\Sigma(t, S_t; K, T)$ are entirely deterministic (although *a priori* not explicit) functions of their four arguments. However, the *sliding* implied volatility surface $\tilde{\Sigma}(t, y, \theta)$ is itself a stochastic function of its three arguments: when the log-moneyness y and time-to-maturity θ are fixed the IV functional becomes *parametrised* by S_t .

Such local volatility models provide an easy understanding of the concept, but they cannot incorporate the notion of *unobservable* state variables. Let us therefore present a more complex illustration, involving a multi-dimensional driver.

Example 2.2 (Markovian dimension of the IV with an independent stochastic volatility) We assume a stochastic instantaneous volatility model with state variables (t, S_t, σ_t) and driven by a bi-dimensional Wiener. Furthermore, we take the dynamics of the volatility σ_t to be purely exogenous as per

$$\begin{cases} \frac{dS_t}{S_t} = \sigma_t dW_t \\ d\sigma_t = f(t, \sigma_t) dt + \vec{h}(t, \sigma_t)^\perp d\vec{Z}_t \end{cases} \quad \text{with} \quad W_t \perp \vec{Z}_t.$$

Then the absolute Call prices $C(t, S_t, \sigma_t; K, T)$ are deterministic functions of their five arguments. In turn, the same is true of the absolute Lognormal implied volatility $\Sigma(t, S_t; \sigma_t; K, T)$.

However, that map can also be seen as a deterministic function of the four market variables $(t, S_t; K, T)$ which has been parametrised by the hidden (or unobservable) process σ_t . In that case the $\Sigma()$ functional becomes stochastic, is driven purely by the exogenous driver $\vec{Z}_t$ and is therefore measurable w.r.t. the latter's filtration. As for the sliding IV surface $\tilde{\Sigma}(t, y, \theta)$, it can be seen as a deterministic function of three variables, parametrised by S_t and σ_t .

In this example, the orthogonality assumption is only there to reinforce the point that the filtration w.r.t. which the sliding “function” $\tilde{\Sigma}$ is measurable is necessarily much finer than the one sufficient to measure its absolute counterpart Σ . In other words, we have incorporated the driver W_t into the sliding quantity.

Having described the relevant mathematical features of the slide, let us now turn to its financial motivation. It revolves mainly around human factors, in particular our limited ability to comprehend high-dimensional and noisy patterns. We expose the practicality and interest of comparing market and model smile dynamics, and how the sliding representation can help in that process.

In our view, one defining characteristic of a well-chosen model is to *minimise re-calibration*, in other words to exhibit stable (or stationary) calibrated parameters. Ideally, we would like to maintain these constant, and explain all joint movements of the underlying and of the smile “through” the model and its drivers. After all, if the market was kind enough to follow known and stationary laws, as is mostly the case in physics or mechanics, that is exactly what would happen. And there is no argument from practitioners that such a (hypothetical) situation is rather more palatable than having to frequently re-adjust these parameters, a procedure that generates additional Mark-To-Market (MTM) and tracking error noise, and therefore hedging (transaction) costs.

However, in order to even get close to such a stationary behaviour, i.e. to be very “realistic”, stochastic instantaneous volatility models generally need to use a significant number of parameters (roughly half a dozen for SABR or for Heston, and even then re-calibration is too frequent). Furthermore, these parameters have very distinct individual impacts on the smile, both in quality and in magnitude. Therefore, attempting to analyse a model's stationarity by observing a collection of historical time series (one for each parameter) might be an interesting academic exercise but *a priori* not a very practical or useful idea.

A much more intuitive approach, in contrast, is to compare the actual market *smile*, at each historical sampling time t_i , to the “prediction” given by the model, itself (statically) calibrated at the previous time t_{i-1} . Obviously the notion of prediction must be made precise, in the sense that this smile's dynamics must be made measurable w.r.t. a given filtration or observable state variables.

In a two-dimensional model such as Heston, SABR, or even some multi-scale extensions,⁷ a simple approach consists in using the spot S_t and a single, very short expiry At-The-Money option (C_t^{IATM}) in order to access the full filtration. Indeed, these market instruments are usually very good proxies for the actual state variables. Using then a Euler approximation, it is typically possible to unequivocally associate driver increments $[\Delta W_t, \Delta Z_t]$ to an historical market movement $[\Delta S_t, \Delta C_t^{IATM}]$. Then one can generate the whole “conditional” smile at t_i , which is to be compared to the actual market smile observed. Repeating this process, it is even possible to compute maximum likelihood estimators, and/or to gauge the descriptive/predictive quality of the model.

Such a comparison of smiles, resulting in a *surface of differences*, is easy to interpret, and can also be visualised in motion, in order to assess the dynamic properties of the calibrated model. But in order to facilitate this interpretation, it is important to choose a common representation (i.e. axis coordinates) that tends to stabilise both smiles, and therefore their difference. In other words, we are looking at a change of coordinates that will make both the market-observed and the model-generated smiles as stationary as possible.

It happens, though, that most market smiles demonstrate obvious and simple sliding properties, both in maturity and in strike. The latter is often referred to as “stickiness” (see [10], for instance, for the sticky-strike or sticky-Delta rule). It also happens that this is actually a property that stochastic volatility models are well suited to capture (see [11], for instance, which provides a good comparison with local volatility models).

In summary, the choice of the new coordinates is certainly motivated by the Black formula itself, and in particular the LOG-moneyness $y = \ln(K/S_t)$ that allows scaling and brings some symmetry to the strike axis. But this choice is also brought forward by a healthy desire for stationarity within a realistic modelling framework, and is in no way binding.

2.1.2.4 Illustration and Limitations

Let us demonstrate how to “cast” into our framework a simplistic model/smile combination.

Example 2.3 (SV Normal model with deterministic rates and carry) We place ourselves in the arch-classical case where X_t is the price process of a traded asset, modelled with a Normal⁸ volatility (which can be local and/or stochastic),

⁷ This is not the case, in particular, for “double mean-reverting” models such as “Double Heston”, which are tri-dimensional: see [8] and [9].

⁸ This is a writing convention, and chosen mainly for demonstration purposes, since it is rather unusual in the Equity world.

and where the short rate $r(t)$ and the dividend/carry rate $d(t)$ are considered deterministic:

$$dX_t = [r(t) - d(t)] dt + \gamma_t dW_t,$$

where W_t is a risk-neutral driver, or $\mathbb{Q}$ -Wiener process.

We simply select the deterministic function $D(t) = \exp(\int_0^t [r(s) - d(s)] ds)$ as our numeraire, which is nothing else than a capitalisation process or money market account. We check that it is a tradeable asset, simply because it is deterministic and therefore can be replicated with liquid assets, the zero-coupons.

Denoting by $S_t \triangleq X_t/D(t)$ the “discounted” asset, and $\sigma_t \triangleq \gamma_t X_t$ the “Lognormal” volatility, we indeed obtain the required dynamics (2.1). As for the option field, the payments are deemed to occur at time T , for an amount of

$$C(K, T) = (X_T - K)^+ = D(T) \left[S_T - \frac{K}{D(T)} \right]^+.$$

It therefore suffices to modify the unit in which we measure the strike, from “cash” K to “discounted” $K/D(t)$, to complete the “cast”. Note that the measure has not changed: it is still the risk-neutral measure $\mathbb{Q}$.

Beyond this trivial example, the chosen framework can fit a wide range of underlyings and options. Of course, it also exhibits several limitations, which prevent the coverage of more complex modelling configurations.

A first restriction is that we defined the setup for a *scalar* driver W_t , which excludes the description of full multi-underlying dynamics. Typical cases of multi-dimensional frameworks occur naturally in the FX or equity environments, with baskets or indexes, for instance.

Its second shortfall is that it is not suitable to deal with a *term structure*, and in particular the case of stochastic rates. Indeed, if one wishes to define a whole smile, then the same numeraire N_t must be invoked in the payoff (2.2), whatever the expiry T . In Example 2.3 we bypassed this issue by choosing a deterministic numeraire, but in the general case of stochastic rates the whole setup must be based on a single maturity T , and must be financially meaningful, it can only deal with the associated implied volatility.

This can be seen as both a special case and an extension of the multi-dimensional framework. Indeed, even under a Black-Scholes model, if the short rate is made stochastic then pricing a Call of maturity T_2 usually entices us to use the forward measure $\mathbb{Q}^{T_2}$ and the associated Zero Coupon $B_t(T_2)$ as numeraire. Should we be only interested in that single expiry, then the problem could be treated in the multi-dimensional setup mentioned above. But then for $T_1 < T_2$ the considered payoff would be $B_{T_1}(T_2) [X_{T_1}/B_{T_1}(T_2) - K]^+$, which is not a liquid product. Therefore the setup would lose its consistency and its financial appeal.

We will therefore extend the current simple setup to cover both these cases, respectively in Sect. 3.4 and in Chap. 5. It will turn out that the multi-dimensional case actually lies in the same conceptual class as the current framework and provides

very similar results, and also that its added difficulty is mainly computational. The term structure extension, however, shares the same technical difficulty as the former, but is structurally on a very distinct level.

2.1.3 The Two Stochastic Volatility Model Frameworks

From now on, we will focus on the Sliding Implied Volatility Surface $\tilde{\Sigma}(t, y, \theta)$ associated to a given model: we are interested in its shape and also in its joint dynamics with the underlying. In our framework, we specifically want this map to exhibit stochastic dynamics, which should be driven by two orthogonal Wiener processes:

- The *endogenous* driver of the underlying, denoted W_t . In full generality the dimension of this driver will be notated as n_w , but initially it will be taken as mono-dimensional since our underlying itself has been defined as a scalar. The consequences of relaxing this assumption will be exposed in Sect. 3.4.
- The *exogenous* driver $\vec{Z}_t$, which enables movements of the implied volatility surface independently of the underlying dynamics. It is taken as multi-dimensional (with finite dimension n_z) to allow for the complex deformation modes observed in practice.

By convention, we will take W_t and $\vec{Z}_t$ to be independent, and all multidimensional Brownian motions (including $\vec{Z}_t$) to be uncorrelated (i.e. to exhibit a *unit* covariance matrix). One might question why we chose to express our dynamics along two uncorrelated Wiener processes. Indeed, other authors have opted for a single, unified driver: this is the case, for instance, in [12]. Clearly this is mathematically insignificant, and purely a matter of presentation. Our view is that it brings two main advantages, for only one drawback.

The first advantage is technical, and is analogous to manipulating independent (as opposed to correlated) Gaussian vectors. The volatility and correlation structures are then combined in (products of) tensorial coefficients, which simplifies the computation of brackets $\langle d\cdot, d\cdot \rangle$.

The second advantage is linked to modelling and interpretation. Indeed, we are attached to the incompleteness, endogenous/exogenous interpretation detailed above. Also, for option pricing purposes and certainly in numerical terms, it makes sense to orthogonalise the drivers.

The single shortfall that we see is also interpretational, in that using our convention we somewhat lose the intuition that drivers' increments are actually "representing" variations in "physical" quantities, such as asset prices, yields or instrumental processes. For instance, the Heston model (see [2]) is traditionally defined as

$$\frac{dX_t}{X_t} = \mu dt + \sqrt{V_t} dW_t \quad (2.6)$$

$$\langle dW_t, dB_t \rangle = \rho dt$$

$$dV_t = \kappa [\theta - V_t] dt + \varepsilon \sqrt{V_t} dB_t. \quad (2.7)$$

But it might make more financial sense to write the correlation structure as

$$\left\langle \frac{dX_t}{X_t}, dV_t \right\rangle = \rho \varepsilon V_t dt \quad \text{rather than} \quad dB_t = \rho dW_t + \sqrt{1 - \rho^2} dZ_t.$$

Going back to our framework, it is clear that one role of $\vec{Z}_t$ is to introduce market incompleteness. It can, in particular, embody *model ambiguity*, since it underscores a finer filtration than the σ -field generated by S_t . As mentioned, it certainly makes it possible for the (sliding) implied volatility surface to move independently of the underlying S_t (i.e. not to be purely *local*) or of its driver W_t (i.e. to exhibit an *exogenous* component). Such a rich behaviour of the $\tilde{\Sigma}(t, y, \theta)$ map can be generated in several ways. In this chapter, we consider only a couple of these distinct model classes: the stochastic *instantaneous* volatility model, and the stochastic *implied* volatility model.⁹

Formally, the specifications of these two model classes (instantaneous SV and implied SV) only share the generic underlying's dynamics (2.1). Indeed, they both describe the shape and the joint dynamics (of the underlying and of the vanilla options) but in different ways; let us now introduce and formalise both these setups.

2.1.3.1 The Generic Stochastic Instantaneous Volatility Model

In this framework, the shape and dynamics of the smile are generated by specifying “in depth” the dynamics of the instantaneous volatility σ_t , using a Wiener chaos representation. Formally, we have to assume a system of SDEs, starting with

$$dS_t/S_t = \sigma_t dW_t \tag{2.8}$$

$$d\sigma_t = a_{1,t}dt + a_{2,t}dW_t + \vec{a}_{3,t}d\vec{Z}_t \quad \text{with} \quad W_t \perp \vec{Z}_t. \tag{2.9}$$

The stochastic coefficients $a_{1,t}$, $a_{2,t}$ and $\vec{a}_{3,t}$ are deemed to be processes, but only imposed to be Markovian and adapted, hence the “generic” denomination for the model. Indeed, most stochastic instantaneous volatility models that are used in practice fall into the *parametric diffusion* category. In that case, the $a_{i,t}$ coefficients are actually parametric functions of a finite collection of state variables, which usually include t , S_t and σ_t . Although the Markovian dimension can get higher than three in this framework, for instance with multi-scale processes,¹⁰ we do not restrict ourselves to such cases. Instead our framework contains those parametric diffusion models, and also gets its “universality” from the fact that the dynamics of $a_{1,t}$, $a_{2,t}$ and $\vec{a}_{3,t}$

⁹ It is also possible to employ stochastic *local* volatility models, see [13].

¹⁰ In general, this implies an increase in the dimension of the driver $\vec{Z}_t$.

are themselves defined by a “chaos” structure. For instance, we specify the (input) dynamics of $a_{2,t}$ in the following way:

$$da_{2,t} = a_{21,t}dt + a_{22,t}dW_t + \vec{a}_{23,t}d\vec{Z}_t. \quad (2.10)$$

However, the dynamics of a multi-dimensional coefficient, such as $\vec{a}_{3,t}$, generate a practical difficulty. Evidently, its dynamics can be similarly symbolised as

$$d\vec{a}_{3,t} = \vec{a}_{31,t}dt + \vec{a}_{32,t}dW_t + \vec{\vec{a}}_{33,t}d\vec{Z}_t,$$

where $\vec{\vec{a}}_{33}$ is an $n_z \times n_z$ matrix. If we continued the specification, then $\vec{\vec{\vec{a}}}_{333,t}$ would be a tensor of order 3, and so on. This is clearly not a promising way to conduct computations, at least by hand. Nevertheless, we will explore this avenue in Sect. 3.4, but with a view towards (computerised) *automation*.

Note finally that in order to simplify notations, the time dependency will often be omitted in the sequel (for instance $a_{2,t}$ will become a_2).

In this generic SInsV framework, we define the *depth* of a coefficient simply by the number of digits forming its index. For instance, coefficient $a_{2,t}$ has depth one, while $a_{231,t}$ has depth three, etc. We will also see that the coefficients can be arranged according to another logic, by “layers”, which naturally appear in the inductive computation of the smile’s asymptotic (IATM) differentials. For example, the first layer will contain σ_t , $a_{1,t}$, $a_{2,t}$, $a_{3,t}$ and $a_{22,t}$ (see Fig. 2.2). Each layer will also be used to designate a group of corresponding smile IATM differentials.

The depth of the *model* itself is defined as the highest depth reached by all the coefficients describing its dynamics. For instance, the model described by (2.8)–(2.10) would have a depth of 2. Since t , S_t , σ_t and the other $a_{i,t}$ coefficients/processes represent the state variables of the model, if the latter’s depth is finite then so is its Markovian dimension.

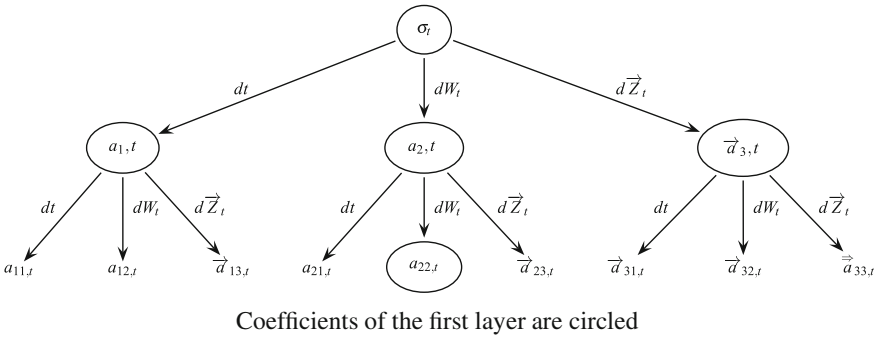


Fig. 2.2 Chaos structure of the generic stochastic instantaneous volatility model

The framework described above, however, is obviously *not* an actual model *per se*. Instead, it consists simply in a *cast*, into which we can arrange any real stochastic instantaneous volatility model, such as the aforementioned parametric diffusions. In practice, most real-life models have a finite and fairly small Markovian dimension (it is 3 for Heston and for SABR) whereas their generic cast can exhibit an infinite number of state variables (infinite depth). We will see, however, that this is not an issue in our asymptotic framework, since—schematically—the higher the depth of a coefficient, the higher the degree of precision it brings to the smile description *via* asymptotic smile differentials.

Therefore, if we consider for instance a SABR model (see [11] and Sect. 4.2 for a description of this local-stochastic volatility model) then the Call price will be a deterministic function of the three Markovian state variables (t , $\vec{S}_t$ and α_t), of the parameter set specifying the diffusion (correlation, vol of vol) and of the option parameters (K and T). Unfortunately, this pricing function is not explicit, as currently only approximations are available (see Sect. 4.2). On the other hand, when casting SABR into our generic framework, we obtain an infinite depth. But the specified dynamics also generate a unique option price surface, by applying no-arbitrage under the chosen measure. This functional cannot *a priori* be expressed either, but we will see that we can instead obtain its asymptotic, potentially infinitely precise description; it simply comes at the cost of an infinite but artificial Markovian dimension.

2.1.3.2 The (Sliding) Stochastic Implied Volatility Model

Modelling the sole dynamics of the implied volatility surface is not a very new idea in itself. A number of empirical studies have been conducted on real data in order to statistically infer the deformation modes of this surface, either in a parametric fashion or not. Usually and rather logically, these empirical investigations have been conducted on very liquid equity indexes, hence minimising the use and influence of interpolation/extrapolation methods.

In [14] the authors analyse the S&P500 and FTSE100 liquid options, with daily frequency. They use a Karhunen-Loève decomposition, which is a generalisation of the Principal Component Analysis to higher dimensional random fields. They uncover, among other interesting features, a typical level/slope/curvature repartition of the leading eigenmodes, as well as characteristic values for the mean-reversion of these modes.

Another interesting presentation can be found in [15], which focuses on semi-parametric modelling, but also covers several inference techniques, as well as practical data processing pitfalls (smoothing in particular).

Modelling the *joint* dynamics of the underlying with the smile, in particular establishing and respecting the structural no-arbitrage constraints, represents a more involved exercise. Apart from stochastic *local* volatility models,¹¹ the main academic

¹¹ See [13]: this class is not to be confused with the local stochastic volatility (LSV) models such as those described in [16], for instance.

attraction has been with stochastic *implied* volatility (SImpV) models. A specificity of this class is to define the initial shape of the smile as an input, and then to model the joint underlying/smile dynamics.

In that respect, there is a strong similitude with the approach that [17] introduced with a market model for the whole yield curve, a.k.a. the HJM framework. Indeed, one could summarily consider that the role of the short rate r_t is now taken by the underlying S_t and its volatility σ_t , while the one-dimensional map of Zero-Coupon prices $T \rightarrow B_t(T)$ is replaced by a bi-dimensional mapping $(K, T) \rightarrow \Sigma(K, T)$. We will see that the resemblance is carried even further, in the sense that σ_t is asymptotically included in the Σ dynamics, in the same way that r_t is embedded in the $B_t(T)$ term structure. We shall first cover some of the papers dealing with the general framework and its structural constraints, before moving on to actual models.

The concept of a SImpV model has been introduced in stages and rather independently by several authors, first for a single option and then for whole smiles. Historically, let us first cite [18] and [19], with the latter using a sliding representation. In [20] the authors present four different versions of the no-arbitrage condition for smile dynamics: for the implied volatility or implied variance, and in absolute or sliding coordinates. Then they apply these results in several practical contexts, including a single Caplet within a BGM model.

Modelling-wise [7] proposes a factor-based instance of the class. This article provides a good interpretation for certain equations, makes a deliberate effort to relate the model to other classes, and contains real-market (DAX) applications. Interestingly, it introduces the notion of a generalised *moneyness*, which can be $\ln(K/S_t)$ but is not restricted to that case: in principle it should be chosen so as to render the *stickiness* of the smile considered.

Another seminal collection is [21–24], which focuses not on the smile but on the variance curve. These articles analyse the intrinsic limitations of SInsV models, such as Heston. They propose several dynamic models for the forward variance, based on a Markovian factor representation, using either a continuous or discrete structure.

In [25] one can also find a focus on volatility derivatives, but also the parametrisation of a specific SImpV model. That class (the *Market Model of Implied Volatility*) exploits *local* dynamics, i.e. a diffusion involving only the following state variables: t , S_t and the smile itself. A particular instance is then proposed (the *Skew market Model*) which models the smile as a parabolic function of log-moneyness.

In essence, stochastic implied volatility models represent the next logical step in the natural evolution of modelling practice, within a given derivative market. Indeed, as more liquid derivative products appear, those need to be included in the calibration. The two main avenues are therefore to complexify the existing models and/or to assume that the new set of products represents an input.

In the matter at hand this fits rather well the recent modelling history. Starting from a situation where only ATM options were liquid, Black-Scholes with deterministic volatility was sufficient. As OTM and ITM options became liquid, the model was complexified, upgraded to instantaneous stochastic volatility. First with local volatility (Dupire), then with stochastic volatility (e.g. Heston, hence increasing the Markovian dimension), and currently with a combination of both (SABR, FL-SV).

As the liquidity of the smile increases, so does the need to describe its dynamics, if only because more and more exotic products depend on it, whether in their definition (volatility derivatives) or for their hedge. The next logical step therefore seems to be the incorporation of that smile within the model, along with its dynamics.

The practical difficulty with that model class, however, resides in the parametrisation of an implied volatility surface that starts and stays valid. By valid we mean that the associated option price surface must satisfy the usual non-arbitrage conditions, everywhere in the (K, T) domain, at any time and almost surely. Arguably, this has been the strongest hurdle in the practical introduction of the stochastic implied volatility model class. However, we believe that this model class may ultimately become just as successful as the HJM framework once was, and also as successful as the LMM¹² framework currently is.

As for our version of the SImpV model, it is defined as follows

$$\begin{cases} dS_t/S_t = \sigma_t dW_t & (2.11) \\ d\tilde{S}(t, y, \theta) = \tilde{b}(t, y, \theta)dt + \tilde{v}(t, y, \theta)dW_t + \vec{\tilde{n}}(t, y, \theta)^\perp d\vec{Z}_t & (2.12) \end{cases}$$

$$\text{with} \quad y \triangleq \ln(K/S_t) \quad \theta \triangleq T - t > 0 \quad W_t \perp \vec{Z}_t$$

where the drivers W_t and $\vec{Z}_t$ are independent¹³ standard Brownian motions under the (chosen) martingale measure $\mathbb{Q}^N$. All dynamic coefficients (σ_t , $\tilde{b}$, $\tilde{v}$ and $\vec{\tilde{n}}$) are taken as generic stochastic processes, so that (2.12) cannot be considered *a priori* as an explicit *diffusion*. However, note that, should these coefficients be specified as deterministic (i.e. local) functions (and irrespective of the drivers' dimensions), then the Markovian dimension of the model would still *a priori* be infinite. Again, this situation is analogous to a bi-dimensional version of the HJM framework, where every point on the yield curve represents a state variable.

At first sight, the fact that no specification is given for σ_t might seem surprising, as it appears to make the model under-determined. But in fact that instantaneous volatility is entirely defined through arbitrage by the implied volatility map $\tilde{S}(t, y, \theta)$, as we will soon establish.

Similarly, note that neither the nature nor the dynamics of the other coefficients $\tilde{b}$, $\tilde{v}$ and $\vec{\tilde{n}}$ are specified. In principle they can be Itô processes also driven by W_t and/or $\vec{Z}_t$, or even by entirely new (and orthogonal) Wiener processes. It happens that, for our intents and purposes, this level of definition is in fact irrelevant, and the justifications for such a simplifying feature will become apparent as the book progresses. In particular, these topics will be discussed in Sect. 2.2.4 and in Chap. 3, which is dedicated to more elaborate versions of the methodology. In summary, the SImpV model (2.11)–(2.12) is well-defined and self-contained as it is.

¹² Libor Market Model, see Chap. 7.

¹³ To lift any remaining ambiguity, this implies that the correlation matrix of $\vec{Z}_t$ is diagonal.

In spirit, moving from a *stochastic instantaneous volatility* model to a *stochastic implied volatility* model is similar to moving from a short-rate model (such as Vasicek) to an HJM model: it can be seen as a simple matter of number of parameters vs number of constraints. Indeed, most short rate models are only capable of generating a certain functional class of yield curves (YC), so that calibration to the bond market is already an issue. In order to recover a given yield curve, one will usually have to add a time-dependent drift, while matching further constraints, such as some marginal distributions of the YC, will require a substantial complexification of the model class: the Hull and White Extended Vasicek [26] comes to mind, which is nothing other than an instance of the HJM class.

On the other hand, opting for an HJM model enables the modeller to calibrate to any yield curve, because this map becomes an integral part of the model. The remaining parameters, or degrees of freedom, are then used to calibrate to the marginal or joint distributions of the YC, to liquid Interest Rates options such as Caps, Swaptions, Bond options, CMS, etc. But as more and more products become liquid, the calibration set tends to increase, so that eventually there are too many constraints and not enough parameters. This is where *stochastic implied volatility* models can come into play: because the liquid Call prices become an integral part of the model, new degrees of freedom become available to calibrate to other (more recent) liquid options.

Although the underlying's instantaneous volatility σ_t is apparently a free parameter of the (sliding) implied volatility model, it is actually to be considered as a formal expression; indeed, we will see that arbitrage constraints impose that σ_t be entirely determined by the stochastic map $\tilde{S}(t, y, \theta)$ (see (2.36), p. 50 and Remark 2.4, p. 51). This precision explains why the dynamics of σ_t are not explicitly specified in the *stochastic implied volatility* model, as was the case with the *stochastic instantaneous volatility* model: indeed they are already included.

2.1.3.3 Comparison, Assumptions and Remarks

The two models share the same dynamics for the underlying, and in particular the Lognormal instantaneous volatility. For the SInsV it is the dynamics of σ_t that are defined in a chaos expansion, while the SImpV specifies the smile dynamics: in other words the SInsV model is defined “in depth” while the SImpV is specified spatially. The choice of a single-dimensional driver W_t for the underlying, as specified respectively by (2.8) and (2.11), is actually benign in both cases and for similar reasons. The only interest in employing a multi-dimensional endogenous driver $\vec{W}_t$ and volatility $\vec{\sigma}_t$ is to describe the joint dynamics of the underlying asset S_t along with another process. In the SImpV case, that process is the smile shape $\tilde{S}$, while for the SInsV model it will be the coefficients in the Wiener chaos decomposition of σ_t : in both cases, these are *a priori* infinite dimensional processes. But in both instances, any component *not* driving can be formally allocated to the independent, exogenous driver $\vec{Z}_t$, so that in the end only the modulus $\|\vec{\sigma}_t\|$ matters. In essence, this all boils down to the very definition of an exogenous noise, whose instantaneous covariation with S_t must be null.

We will however work with a multi-dimensional $\vec{W}_t$ later on, in Chap. 3, in order to investigate the basket problem. In that context, we will have to express the joint dynamics of each individual underlying and of the basket itself, which does warrant a vectorial endogenous driver.

In terms of presentation, the fact that we organised the drivers into orthogonal components is clearly artificial, but not binding. In purely economic terms, there is no such thing as clearly identifiable independent factors. It serves several purposes. In mathematical terms, it simplifies the computations and ensures that all correlation-related quantities are represented by linear algebra products. As for interpretation, it clearly divides the picture between an endogenous/observable/complete component, and the exogenous/non-observable/incomplete part. Obviously there is a cost to pay: in particular the bracket between two processes now comes as a scalar product. For instance the notion associated to $\langle \frac{dS_t}{S_t}, d\sigma_t \rangle$ is easy to grasp, even to graphically chart on historical series, whereas $a_{2,t}$ might initially seem a bit abstract. Nevertheless, the logic used throughout this book is to clarify (even artificially) the computation, and re-formulate the output results for interpretation.

Remark 2.1 The choice of the Lognormal convention to write the dynamics of the underlying S_t (see (2.8) and (2.11)), as well as the fact that we are re-parameterising the price surface using a Lognormal (Black) convention, might seem subjective and possibly restrictive. In fact, as will be proven and discussed in Sect. 3.3.3, p. 164, once expansion results are available for a simple model such as the Lognormal dynamics, it is relatively simple to transfer those to most parametrisations, such as the Normal dynamics, or the CEV, etc. In consequence the choice we made is merely practical, and in practice not binding.

In order to facilitate the coming proofs, we add the following technical restrictions:

Assumption 2.1

$$\text{Almost surely} \quad S_t > 0 \quad \forall t \geq 0 \quad (2.13)$$

$$\text{Almost surely} \quad \sigma_t > 0 \quad \forall t \geq 0 \quad (2.14)$$

$$\text{Almost surely} \quad \tilde{\Sigma}(t, y, \theta) > 0 \quad \forall (t, y, \theta) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \quad (2.15)$$

Remark 2.2 In practice, Assumption 2.14 is not as restrictive as it might seem. As will become apparent in Sects. 2.3 and 3.4, the positivity of σ_t is in fact equivalent to the positivity of $\tilde{\Sigma}(t, 0, 0)$, which in a broader multi-dimensional context, is equivalent to the positivity of the modulus $\|\vec{\sigma}_t\|$. Therefore, in such a framework, one could see any of the components of $\vec{\sigma}_t$ go null: as long as one component remains either strictly positive or negative, the following computations are valid.

In fact, because our approach is asymptotic, Assumption 2.13 needs only hold at the current time t . We could still express results should it be breached, but they would be trivial and without financial interest.

2.1.4 The Objectives

The objectives of this chapter are to *establish the links between the stochastic (sliding) implied volatility model and the instantaneous volatility model*.

- The *direct problem* is to derive the shape and dynamics of the implied volatility surface $\tilde{\Sigma}$ from the value and dynamics of the instantaneous volatility σ_t .
- The *inverse problem* is to derive the value and dynamics of the instantaneous volatility σ_t from the shape and dynamics of the implied volatility surface $\tilde{\Sigma}$.

The (static) *calibration procedure* is often viewed as an inverse problem. Indeed, when a practitioner exploits one of the many popular stochastic instantaneous volatility models, the implied volatility $\tilde{\Sigma}$ must be marked to market in some respect, while σ_t is intrinsically model-dependent. In numerical terms, however, this inverse problem is usually solved by an optimisation process. The latter consists in minimising the market error, which itself requires numerous calls to the pricer, and is thus associated to the *direct* problem.

Let us therefore begin by examining the structural constraints of the SImpV model.

2.2 Derivation of the Zero-Drift Conditions

The stochastic implied volatility model, as defined by (2.11)–(2.12), is just an SDE system. It describes the dynamics of an infinite-dimensional state vector (the underlying S_t and the smile $\tilde{\Sigma}(t, y, \theta)$) with no built-in notion of how they relate to each other, hence it does not intrinsically guarantee no-arbitrage, a condition that must be imposed externally.

We start by establishing the main Zero-Drift Condition, which is valid in the full domain (t, y, θ) . Then we specialise it to the Immediate domain, i.e. $\theta = 0$, which provides a pair of Immediate ZDCs. Finally, we restrict ourselves to the Immediate ATM position, which is the starting point of our asymptotics.

2.2.1 The Main Zero-Drift Condition

Let us first transfer the dynamics of the sliding smile back into absolute coordinates.

Lemma 2.1 (Dynamics of the absolute implied volatility surface) *In our framework, the dynamics of the absolute implied volatility are*

$$d\Sigma(t, S_t, K, T) = b(\alpha) dt + v(\alpha) dW_t + \vec{n}(\alpha)^\perp d\vec{Z}_t \quad (2.16)$$

where the coefficients are given by

$$\begin{aligned} b(\infty) &= \tilde{b}(\circ) - \tilde{\Sigma}'_{\theta}(\circ) + \frac{1}{2} \sigma_t^2 \left[\tilde{\Sigma}''_{yy}(\circ) + \tilde{\Sigma}'_y(\circ) \right] - \sigma_t \tilde{v}'_y(\circ) \\ v(\infty) &= \tilde{v}(\circ) - \sigma_t \tilde{\Sigma}'_y(\circ) \\ \vec{n}(\infty) &= \vec{n}(\circ) \end{aligned} \quad (2.17)$$

with absolute and sliding arguments defined as $(\circ) \triangleq (t, y, \theta)$ and $(\infty) \triangleq (t, S_t, K, T)$.

Proof Let us invoke the Itô-Kunita formula as in Theorem A.1. Taking $\vec{\alpha}_t \triangleq \begin{pmatrix} y \\ \theta \end{pmatrix}$ leads to

$$dy = -\frac{1}{S_t} dS_t + \frac{1}{2} \frac{1}{S_t^2} \langle dS_t \rangle = \frac{1}{2} \sigma_t^2 dt - \sigma_t dW_t \quad \text{and} \quad \langle dy \rangle = \sigma_t^2 dt.$$

Therefore the dynamics of the absolute $\Sigma(t, S_t, K, T)$ surface are

$$\begin{aligned} d\Sigma &= \tilde{b}(\circ)dt + \tilde{v}(\circ)dW_t + \vec{n}(\circ)^\perp d\vec{Z}_t + \tilde{\Sigma}'_y(\circ)dy + \tilde{\Sigma}'_{\theta}(\circ)d\theta \\ &\quad + \frac{1}{2} \tilde{\Sigma}''_{yy}(\circ) \langle dy \rangle - \tilde{v}'_y(\circ) \sigma_t dt \\ &= \tilde{b}(\circ)dt + \tilde{v}(\circ)dW_t + \vec{n}(\circ)^\perp d\vec{Z}_t + \tilde{\Sigma}'_y(\circ) \left[\frac{1}{2} \sigma_t^2 dt - \sigma_t dW_t \right] \\ &\quad - \tilde{\Sigma}'_{\theta}(\circ)dt + \frac{1}{2} \tilde{\Sigma}''_{yy}(\circ) \sigma_t^2 dt - \tilde{v}'_y(\circ) \sigma_t dt. \end{aligned}$$

Then grouping the finite and non-finite variation terms provides the desired result. $\square$

Having now moved to an absolute setup enables us to use the martingale property, and therefore to express our main result.

Proposition 2.1 (Zero Drift Condition for a single underlying) *The shape and dynamics functionals of the sliding SImpV model (2.11)–(2.12) are constrained by arbitrage to fulfil the following **Zero-Drift Condition**:*

In the general domain $(\circ) = (t, y, \theta) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^{+}$ we have a.s.*

$$\tilde{\Sigma}^3(\circ) \tilde{b}(\circ) = \theta D(\circ) + E(\circ) + \frac{1}{\theta} F(\circ) \quad (2.18)$$

with

$$D(\circ) = \frac{1}{8} \tilde{\Sigma}^4(\circ) \left[\left[\tilde{v}(\circ) - \sigma_t \tilde{\Sigma}'_y(\circ) \right]^2 + \|\vec{n}\|^2(\circ) \right] \quad (2.19)$$

$$E(o) = \tilde{\Sigma}^3(o) \left[\tilde{\Sigma}'_{\theta}(o) - \frac{1}{2} \sigma_t^2 \tilde{\Sigma}''_{yy}(o) + \sigma_t \tilde{v}'_y(o) - \frac{1}{2} \sigma_t \tilde{v}(o) \right] \quad (2.20)$$

$$\begin{aligned} F(o) = & \frac{1}{2} \tilde{\Sigma}^4(o) - \frac{1}{2} \sigma_t^2 \tilde{\Sigma}^2(o) - y \sigma_t \tilde{\Sigma}(o) \left[\tilde{v}(o) - \sigma_t \tilde{\Sigma}'_y(o) \right] \\ & - \frac{1}{2} y^2 \left[\left[\tilde{v}(o) - \sigma_t \tilde{\Sigma}'_y(o) \right]^2 + \|\vec{n}\|^2(o) \right]. \end{aligned} \quad (2.21)$$

Proof In order to obtain the Call price dynamics, we start by applying Itô's Lemma to the normalised Black-Scholes functional (2.3). For the sake of clarity, we omit the multiple dependencies of the implied volatility Σ , of the normalised Greeks Δ ("Delta"), Γ ("Gamma"), $\mathcal{V}$ ("Vega"), Λ ("Volga"), ϑ ("Vanna") and of the absolute diffusion coefficients b, v, n . We obtain simply:

$$\begin{aligned} dC(t, S_t, T, K) = & -\mathcal{V} \Sigma (2\sqrt{\theta})^{-1} dt + \Delta S_t \sigma_t dW_t \\ & + \mathcal{V} \sqrt{\theta} \left[b dt + v dW_t + \vec{n}^\perp d\vec{Z}_t \right] \\ & + \frac{1}{2} \Gamma S_t^2 \sigma_t^2 dt + \frac{1}{2} \theta \vartheta \left(v^2 + \|\vec{n}\|^2 \right) dt + \Lambda \sqrt{\theta} S_t \sigma_t v dt \end{aligned} \quad (2.22)$$

with

$$\Delta = \frac{\partial C^{BS}}{\partial x} \quad \mathcal{V} = \frac{\partial C^{BS}}{\partial v} \quad \Gamma = \frac{\partial^2 C^{BS}}{\partial x^2} \quad \Lambda = \frac{\partial^2 C^{BS}}{\partial x \partial v} \quad \vartheta = \frac{\partial^2 C^{BS}}{\partial v^2}.$$

The No-Arbitrage Assumption forces C_t into a martingale under $\mathbb{Q}^N$, so that (2.22) leads classically to the following zero-drift condition:

$$\begin{aligned} 0 = & -\mathcal{V} \Sigma (2\sqrt{\theta})^{-1} + \mathcal{V} \sqrt{\theta} b + \frac{1}{2} \Gamma S_t^2 \sigma_t^2 \\ & + \frac{1}{2} \theta \vartheta \left(v^2 + \|\vec{n}\|^2 \right) + \Lambda \sqrt{\theta} S_t \sigma_t v. \end{aligned} \quad (2.23)$$

Computing the normalised Greeks $\mathcal{V}$, Γ , ϑ and Λ involved in (2.23) presents no difficulty, as detailed in Appendix C. Factorising with the Vega, the resulting expressions are given by

$$\begin{aligned} \mathcal{V} = S_t \mathcal{N}'(d_1) & \quad \vartheta = \mathcal{V} \left[\frac{y^2}{\left(\Sigma \sqrt{\theta} \right)^3} - \frac{1}{4} \Sigma \sqrt{\theta} \right] \\ \Gamma = \frac{\mathcal{V}}{S_t^2 \Sigma \sqrt{\theta}} & \quad \Lambda = S_t^{-1} \mathcal{V} \left[\frac{1}{2} + \frac{y}{\Sigma^2 \theta} \right] \end{aligned}$$

Note that these expressions are well-defined and finite, thanks to the technical assumptions (2.13) and (2.15). Note also that, although the normalised Delta and Gamma correspond to the same expressions as for the classical BS formula, the derivatives in v (namely Vega, Vanna and Volga) differ slightly from their classical expressions. Note finally that here θ is NOT the Greek associated to time t !

Substituting the Greeks, the zero-drift condition becomes

$$0 = \mathcal{V} \left[-\frac{\Sigma}{2\sqrt{\theta}} + \sqrt{\theta}b + \frac{1}{2} \frac{1}{\Sigma\sqrt{\theta}} \sigma_t^2 \right. \\ \left. + \frac{1}{2}\theta \left[\frac{y^2}{\Sigma^3\theta^{\frac{3}{2}}} - \frac{1}{4}\Sigma\sqrt{\theta} \right] (v^2 + \|\vec{n}\|^2) + \left[\frac{1}{2} + \frac{y}{\Sigma^2\theta} \right] \sqrt{\theta}\sigma_tv \right].$$

Using the strict positivity of $\mathcal{V}$ and of θ (since we are restricted to $\Theta \in \mathbb{R}^{+*}$) we divide both sides by the product $\mathcal{V}\sqrt{\theta}$, then isolate $b(t, y, \theta)$ and end up with

$$b = \frac{1}{2} \frac{\Sigma}{\theta} - \frac{1}{2} \frac{1}{\Sigma\theta} \sigma_t^2 - \frac{1}{2} \left[\frac{y^2}{\Sigma^3\theta} - \frac{1}{4}\Sigma\theta \right] (v^2 + \|\vec{n}\|^2) - \left[\frac{1}{2} + \frac{y}{\Sigma^2\theta} \right] \sigma_tv.$$

We can now use Lemma 2.1 and replace the absolute processes with their sliding counterparts. We get the sliding drift as follows, where all processes are evaluated in (t, y, θ) and with terms gathered by power of θ :

$$\tilde{b} = \theta \left[\frac{1}{8} \Sigma \left[(\tilde{v} - \sigma_t \tilde{\Sigma}'_y)^2 + \|\vec{n}\|^2 \right] \right. \\ \left. + \left[\tilde{\Sigma}'_\theta - \frac{1}{2} \sigma_t^2 (\tilde{\Sigma}''_{yy} + \tilde{\Sigma}'_y) + \sigma_t \tilde{v}'_y - \frac{1}{2} \sigma_t (\tilde{v} - \sigma_t \tilde{\Sigma}'_y) \right] \right. \\ \left. + \frac{1}{\theta} \left[\frac{1}{2} \Sigma - \frac{1}{2} \frac{\sigma_t^2}{\Sigma} - \frac{1}{2} \frac{y^2}{\Sigma^3} \left[(\tilde{v} - \sigma_t \tilde{\Sigma}'_y)^2 + \|\vec{n}\|^2 \right] - \frac{y}{\Sigma^2} \sigma_t (\tilde{v} - \sigma_t \tilde{\Sigma}'_y) \right] \right].$$

Finally, simplifying the second bracket and multiplying the whole expression by $\tilde{\Sigma}^3(t, y, \theta)$ provides the desired expression (2.18), which concludes the proof. $\square$

The ZDC (2.18) links the four parametric processes—all bi-dimensional—describing the sliding IV: the shape $\tilde{\Sigma}$ and the dynamic coefficients $\tilde{b}$, $\tilde{v}$ and $\vec{n}$. We emphasise that this relationship is valid in the full domain (y, θ) as opposed to most of the coming *asymptotic* results, which are either Immediate ($\theta = 0$) or IATM ($y = 0, \theta = 0$). We note the positivity of the highest θ -order term $D(t, y, \theta)$, a property which will prove useful in the sequel. The recurrence of the term

$$[\tilde{v} - \sigma_t \tilde{\Sigma}'_y]$$

is also remarkable, although this term is no stranger to us. Indeed, through (2.17) it is identified as

$$v(t, S_t, K, T)$$

which is the endogenous volatility of the *absolute* stochastic IV surface. In other words, the compensation term

$$\sigma_t \tilde{\Sigma}'_y(\circ)$$

appears to neutralise the space slide associated with movements of the underlying S_t . It would be wrong, however, to assume that it removes *all* dependency on S_t , i.e. that it corresponds to some *unconditional* endogenous volatility of the smile. Indeed, the absolute coefficient v itself can very well incorporate a local component. We note then that the term

$$\left[\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right]^2 + \|\vec{n}\|^2$$

which appears in $D(\circ)$ and $F(\circ)$ represents the *quadratic variation* of the stochastic *absolute* IV surface $\Sigma(t, S_t, K, T)$, and is linked to its *sliding* counterpart with

$$v^2 + \|\vec{n}\|^2 = \left[\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right]^2 + \|\vec{n}\|^2 = \tilde{v}^2 + \|\vec{n}\|^2 - \sigma_t \tilde{\Sigma}'_y \left[2\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right].$$

In that expression, we will see that IATM (i.e. with $y = 0$ and $\theta = 0$) the last term converges to a *negative* correction (see (2.51) [p. 62]). We cannot stress enough that the stochastic PDE (2.18) constitutes the actual basis for most of the subsequent asymptotic expressions: manipulating the ZDC (i.e. differentiating w.r.t. y and θ) then imposing some regularity assumptions, and finally taking the limits in $(t, 0, 0)$, forms the backbone of the ACE methodology.

An interesting insight into the ZDC structure can be given by exploiting the concept of *local volatility* (LV) as per [27]. Indeed, the LV surface $f_{t,S_t}(K, T)$ can be seen as a re-parametrisation of the absolute IV surface $\Sigma_{t,S_t}(K, T)$ via an auxiliary diffusion process. But its square is also interpreted as the expectation of the instantaneous variance σ_t^2 at a given future date T , conditional on the underlying's value ($S_T = K$). In this context we can switch seamlessly to the sliding coordinates y and θ , defining the equivalent *sliding local volatility* as

$$\tilde{f}(t, y, \theta) \triangleq f_{t,S_t}(K, T) \quad \text{so that} \quad \tilde{f}^2(t, y, \theta) = \mathbb{E}_t \left[\sigma_{t+\theta}^2 | S_{t+\theta} = S_t e^y \right],$$

which allows us to naturally isolate the *relative* local variance $\xi_t(t, y, \theta)$ with

$$\xi_t(t, y, \theta) \triangleq \tilde{f}^2(t, y, \theta) - \sigma_t^2 = \mathbb{E}_t \left[\int_t^{t+\theta} d\sigma_u^2 | S_{t+\theta} = S_t e^y \right].$$

We now get slightly ahead of ourselves, by introducing a topic which will be covered in more detail later in this chapter (see p. 66). Indeed, we invoke the classical re-parametrisation of Dupire's formula in terms of Lognormal implied volatility (which can be found in [28]):

$$\tilde{f}^2(t, y, \theta) = \frac{\tilde{\Sigma}^4(\circ) + 2\theta \tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta}(\circ)}{\left[\tilde{\Sigma}(\circ) - y \tilde{\Sigma}'_y(\circ) \right]^2 - \frac{1}{4}\theta^2 \tilde{\Sigma}^4 \tilde{\Sigma}'^2_y(\circ) + \theta \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy}(\circ)},$$

where $(\circ) = (t, y, \theta)$ still denotes the whole domain. First this expression rewrites as

$$\begin{aligned} \tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta}(\circ) &= -\frac{1}{8}\theta \tilde{f}^2 \tilde{\Sigma}^4 \tilde{\Sigma}'^2_y(\circ) + \frac{1}{2}\tilde{f}^2 \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy}(\circ) \\ &\quad + \frac{1}{\theta} \frac{1}{2} \tilde{f}^2 \left[\tilde{\Sigma} - y \tilde{\Sigma}'_y \right]^2(\circ) - \frac{1}{\theta} \frac{1}{2} \tilde{\Sigma}^4(\circ). \end{aligned}$$

We identify the l.h.s. as the first component of the term $E(t, y, \theta)$ in (2.26). Hence substituting the r.h.s. into the main ZDC (2.18) leads to

$$\tilde{\Sigma}^3 \tilde{b}(\circ) = \theta D^*(\circ) + E^*(\circ) + \frac{1}{\theta} F^*(\circ) \quad (2.24)$$

with

$$D^*(\circ) = \frac{1}{8} \tilde{\Sigma}^4 G^*(\circ) \quad (2.25)$$

$$E^*(\circ) = \tilde{\Sigma}^3(\circ) \left[\frac{1}{2} \xi_t \tilde{\Sigma}''_{yy}(\circ) + \sigma_t \tilde{v}'_y(\circ) - \frac{1}{2} \sigma_t \tilde{v}(\circ) \right] \quad (2.26)$$

$$F^*(\circ) = \frac{1}{2} \xi_t \tilde{\Sigma}^2(\circ) - y \tilde{\Sigma}(\circ) \left[\sigma_t \tilde{v}(\circ) + \xi_t \tilde{\Sigma}'_y(\circ) \right] - \frac{1}{2} y^2 G^*(\circ) \quad (2.27)$$

where

$$G^*(\circ) \triangleq \tilde{v} \left[\tilde{v} - 2\sigma_t \tilde{\Sigma}'_y \right](\circ) - \xi_t \tilde{\Sigma}'^2_y(\circ) + \|\vec{n}\|^2(\circ),$$

which emphasises the pivotal role played by the relative local variance $\xi_t(t, y, \theta)$.

2.2.2 The Immediate Zero Drift Conditions

Let us now focus on the *immediate smile*, which is the limit process (assuming its existence) of the implied volatility surface when time-to-maturity θ tends to 0. In other words, or rather in market terms, the limit of $\tilde{\Sigma}(t, y, \theta)$ when $\theta \searrow 0$ is the implied volatility of a Call option maturing tomorrow, or even in a few hours.

In this context, the assumption of an option continuum (see Sect. 2.1.2.1 [p. 25]) is vital, but it also brings some practical issues. Indeed, *interpolating* between existing Call prices (in order to “fill” the map) is one thing, but *extrapolating* from the liquid Call with the shortest expiry, down to $\theta = 0$, is another modelling and technical problem altogether.

Besides, beyond the continuum and extrapolation hypothesis, we clearly need to assume some additional regularity, in order to ensure the *existence* of a limit. Effectively the ZDC has been established for strictly positive time-to-maturities only, i.e. $(t, y, \theta) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^{+*}$. This was due to the use of Black’s formula within the proof, as a medium between the price functional (necessarily static in $\theta = 0$, since it identifies with the payoff function) and the implied volatility (which can *a priori* afford some dynamics in the “immediate” area $\theta = 0$). Now we need to somehow extend the ZDC into this asymptotic area, which is why we introduce the following sufficient (strong) conditions:

Assumption 2.2 (*Immediate regularity*) Each process $\tilde{\Sigma}$, $\tilde{\Sigma}'_y$, $\tilde{\Sigma}''_{yy}$, $\tilde{\Sigma}'_\theta$, $\tilde{b}$, $\tilde{v}$, $\tilde{v}'_y$ and $\tilde{n}$ admits its own finite (stochastic) limit when $\theta \searrow 0$. These limits are (abusively) denoted with argument $(t, y, 0)$.

If this assumption *package* might appear blunt in the mathematical sense, this is not the case in modelling (i.e. financial) terms. It is true that implied and realised volatilities tend to “pick up” shortly before expiry, often due to the *pinning* effect on very liquid maturities/strikes (see [29] for instance). But to our knowledge such behaviour does not warrant any assumption of instability or of explosion at expiry, as is observed with deltas when a barrier option knocks in/out. For all intents and purposes, at vanishing maturities all our smiles can realistically be considered smooth and well-behaved, in their statics and in their dynamics.

Corollary 2.1 (The Immediate Zero-Drift Conditions) *As a consequence of the ZDC (2.18), shape and dynamics functionals of the sliding SimpV model (2.11)–(2.12) are constrained by arbitrage to respect the following two equivalent Immediate Zero-Drift Conditions (IZDCs):*

In the Immediate domain $(\bullet) = (t, y, 0)$ we have a.s.

- *The Primary IZDC:*

$$0 = F(t, y, 0) = \tilde{\Sigma}^4(\bullet) - \sigma_t^2 \tilde{\Sigma}^2(\bullet) - y 2\sigma_t \tilde{\Sigma}(\bullet) \left[\tilde{v}(\bullet) - \sigma_t \tilde{\Sigma}'_y(\bullet) \right] - y^2 \left[\|\vec{\tilde{n}}(\bullet)\|^2 + \left[\tilde{v}(\bullet) - \sigma_t \tilde{\Sigma}'_y(\bullet) \right]^2 \right]. \quad (2.28)$$

- *The Secondary IZDC:*

$$\tilde{\Sigma}^3 \tilde{b}(t, y, 0) = E(t, y, 0) + F'_\theta(t, y, 0). \quad (2.29)$$

Proof Assumption 2.2 applied to the ZDC (2.18) implies that both its l.h.s. and r.h.s. admit a finite limit. Then considering the last term $\theta^{-1}F(t, y, \theta)$ alone leads to

$$\lim_{\theta \searrow 0} F(t, y, \theta) = 0, \quad (2.30)$$

which proves the Primary IZDC (2.28). Let us now compute the limit of $\tilde{b}(t, y, \theta)$ when $\theta \searrow 0$. Using a small- θ expansion on $F(t, y, \theta)$ we get that

$$F(t, y, \theta) = F(t, y, 0) + \theta F'_\theta(t, y, 0) + O(\theta^2). \quad (2.31)$$

In light of the Primary IZDC (2.28), and since *a priori* $F'_\theta(t, y, 0)$ is non-null, L'Hopital's rule grants us therefore that

$$\lim_{\theta \searrow 0} \frac{1}{\theta} F(t, y, \theta) = F'_\theta(t, y, 0), \quad (2.32)$$

so that, still in the general domain (t, y, θ) , the ZDC (2.18) rewrites as:

$$\tilde{\Sigma}^3 \tilde{b}(t, y, \theta) = \theta D(t, y, \theta) + E(t, y, \theta) + F'_\theta(t, y, 0) + O(\theta). \quad (2.33)$$

Invoking now the regularity package of Assumption 2.2, we can take the limit of (2.33) in $\theta = 0$, which provides the Secondary IZDC (2.29). $\square$

Remark 2.3 (The “true” IZDC) Note that the Secondary Immediate ZDC (2.29) can be considered the legitimate heir of the main ZDC (2.18) in the immediate domain $(t, y, 0)$, whereas what we call the Primary IZDC (2.28) is simply induced by the regularity assumptions.

Nevertheless, for our purposes the latter is simultaneously more practical and more intuitive, as it involves a smaller group of functionals. Indeed, we can observe that in (2.28) the processes $\tilde{b}$, $\tilde{\Sigma}'_\theta$, $\tilde{\Sigma}''_{yy}$ and $\tilde{v}'_y$ play no direct role, by opposition to the general ZDC (2.18) and the Secondary Immediate ZDC (2.29).

Consequently, in the sequel the denomination IZDC will always refer, unless otherwise specified, to the Primary version (2.28).

Interestingly, the limit result (2.30) also provides us with a lower bound for the convergence speed of $F(t, y, \theta)$, which is faster than θ as that time-to-maturity tends to zero. Note, however, that this statement is conditional on the regularity Assumption 2.2, which provides the *existence* of a finite limit for $F(t, y, \theta)$ in the first place.

Looking at the IZDC (2.28) we note that, should the sliding smile $\tilde{\Sigma}(t, y, \theta)$, in the Immediate domain, either be static (in y and θ) or exhibit uniform volatilities $\tilde{v}(\bullet)$ and $\tilde{n}(\bullet)$, then the IZDC would be consistent with a parabolic expression for the Immediate Smile. Indeed, (2.28) can be rewritten as

$$\tilde{\Sigma}^4(\bullet) = \underbrace{\left[\sigma_t \tilde{\Sigma}(\bullet) + y \left[\tilde{v}(\bullet) - \sigma_t \tilde{\Sigma}'_y(\bullet) \right] \right]^2}_{\text{Endogenous}} + \underbrace{y^2 \|\vec{n}(\bullet)\|^2}_{\text{Exogenous}}. \quad (2.34)$$

In our view, this expression stresses the role of the exogenous volatility $\vec{n}()$ as a generator of pure smile convexity. Indeed, there is no interaction with a y term, therefore there cannot be any influence on the skew. Recall also from (2.17) that we can simplify the term $[\tilde{v}(\bullet) - \sigma_t \tilde{\Sigma}'_y(\bullet)]$ into the absolute endogenous coefficient $v(t, S_t, K, T = t)$.

Hence (2.34) can be seen as a “Pythagorean” relationship: in the Immediate domain, the square of the implied “variance” $\tilde{\Sigma}^2$ is the sum of two quadratic and orthogonal terms: one endogenous, the other exogenous. Note also that log-moneyness y acts as “leverage” so that the endogenous term in (2.34) can be interpreted as the endogenous move created at $y = 0$ (i.e. IATM) plus a “torque” proportional to y .

The information provided by the IZDC through (2.34) is actually richer. For simplicity’s sake, let us restrict ourselves to purely endogenous models, which include (but are not restricted to) local volatility models. Given the regularity of all process involved, and in particular assuming that (2.15) extends to the immediate domain, this implies in turn that the sign of the bracket above must stay positive:

$$\tilde{\Sigma}^2(\bullet) = \sigma_t \tilde{\Sigma}(\bullet) + y \left[\tilde{v}(\bullet) - \sigma_t \tilde{\Sigma}'_y(\bullet) \right] \geq 0.$$

We can then divide both sides by $\sigma_t \tilde{\Sigma}^2(t, y, 0)$ to obtain

$$\frac{1}{\sigma_t} = \left[\frac{1}{\tilde{\Sigma}(\bullet)} - y \frac{\tilde{\Sigma}'_y(\bullet)}{\tilde{\Sigma}^2(\bullet)} \right] + y \frac{\tilde{v}(\bullet)}{\sigma_t \tilde{\Sigma}^2(\bullet)} = \left[\frac{y}{\tilde{\Sigma}(\bullet)} \right]'_y + y \frac{\tilde{v}(\bullet)}{\sigma_t \tilde{\Sigma}^2(\bullet)}. \quad (2.35)$$

It is interesting to compare the above expression with similar results in the pure local volatility (LV) case. For instance—and again getting ahead of ourselves—[30] provides the Immediate implied volatility as the *harmonic mean* of the local volatility, over the $[S_t, K]$ segment (see (2.62) [p. 67]). Hence by integrating (2.35) we can interpret the bracket on the r.h.s. as some *local* component.

2.2.3 The IATM Identity

Having expressed the ZDC in the general domain, and then specialised that no-arbitrage constraint to the Immediate domain, it is time to focus further and quote the well-known relationship between instantaneous and implied volatility when taken at the IATM point.

Corollary 2.2 (The IATM Identity) *As a consequence of the IZDC (2.28) and hence by arbitrage:*

At the Immediate ATM (IATM) point $(\star) \triangleq (t, y = 0, \theta = 0)$ we have a.s.

$$\sigma_t = \tilde{\Sigma}(t, 0, 0). \quad (2.36)$$

Proof We simply take the Immediate Zero-Drift Condition (2.28) at the origin point $(t, 0, 0)$, which immediately gives us the fundamental identity. $\square$

The fundamental identity (2.36) in itself is not new, although it has usually been stated in the case of the *absolute* implied volatility surface: see [12], for instance. In practice, this feature of any SInSV model is commonly used in order to “complete the market”, under the assumption that the exogenous driver Z_t is scalar.

In more general terms, hedging stochastic instantaneous volatility can be achieved by incorporating a well-chosen European option in the replication portfolio (refer to [31], for instance). In that respect, selecting a short-dated ATM option often makes sense, for liquidity reasons. Indeed, the At-The-Money Call maturing the soonest tend to be very liquid, if not the most liquid, within the whole price surface. Trading this Call is the natural proxy for trading $\tilde{\Sigma}(t, 0, 0)$ and therefore hedging σ_t .

Another way to comprehend the IATM identity (2.36) is to consider the Gamma-Theta trading of that same option: the implied volatility determines the time decay, a.k.a. Theta, while the instantaneous volatility conditions the Gamma. Therefore, should we not have (2.36), a very obvious (and exploitable) arbitrage opportunity would arise.

Remark 2.4 (Minimal specification of the SImpV model) A consequence of the fundamental result (2.36) is that the (sliding) implied volatility model presented in Sect. 2.1 can be rewritten in a sparser way. Indeed, denoting the generic point by $(\circ) = (t, y, \theta)$ we have

$$\begin{cases} \frac{dS_t}{S_t} = \tilde{\Sigma}(\mathbf{t}, \mathbf{0}, \mathbf{0})dW_t \\ d\tilde{\Sigma}(\circ) = \tilde{\Sigma}^{-3} \left[\theta D + E + \frac{1}{\theta} F \right](\circ)dt + \tilde{\mathbf{v}}(\circ)dW_t + \vec{n}(\circ)^\perp d\vec{Z}_t. \end{cases}$$

In other words, the inclusion of the stochastic instantaneous volatility σ_t within the definition of the sliding implied volatility model is redundant. Indeed, the specification of the sole stochastic map $\tilde{\Sigma}(t, y, \theta)$ formally includes its asymptotics, provided that we assume finite limits in $\theta = 0$. Therefore the statics and dynamics of the V surface entirely determine the SImpV model.

This configuration shows (again) strong similarities with the HJM framework, where the dynamics of the Zero-Coupon are

$$\frac{dB_t(T)}{B_t(T)} = r_t dt + \vec{F}_t(T) d\vec{W}_t$$

and where the drift coefficient r_t , which is the short rate, is itself already (asymptotically) included in the curve input:

$$r_t = -\lim_{T \searrow t} \partial_T \ln [B_t(T)].$$

Also, it is worth recalling that the above interest rates result is, likewise, the structural consequence of no-arbitrage constraints, rather than modelling choices.

2.2.4 *Synthesis and Overture*

To summarise, we have so far expressed the ZDC (2.18) which, invoking regularity assumptions, we have then *specialised* to increasingly restrictive asymptotic domains: first in $\theta = 0$ with the IZDCs (2.28) and (2.29), then at the IATM point with the IATM identity (2.36).

A natural comment would be that other asymptotic domains deserve to be explored, in particular *extreme strikes* (as in $y = \pm\infty$), a topic which is covered for instance in [32, 33] or [34]. Although we have investigated this subject, this book will address it only from a numerical perspective, in Chap. 4.

Going back to the ZDC and its asymptotic corollaries, let us recall that our objective is to solve the direct and/or the inverse problem. In that respect we observe that the ZDC and Immediate ZDCs, while carrying more information (being valid in wider domains) than the IATM identity, seem to provide little information w.r.t. either problem. By contrast, based on the IATM Identity (2.36) and on Remark 2.4, it is tempting to assume that the SInsV model is actually *embedded* into the apparently richer SImpV framework. Indeed, the dynamics of the SImpV smile are specified in every individual point of the (y, θ) map, and can therefore describe complex deformation modes. If such was the case, then the inverse problem would become trivial, while the direct one would become definitively ill-posed.

In fact, this unilateral embedding is verified in certain conditions, dependent in particular on a low dimensionality. But we shall prove that, when it occurs, this inclusion is part of a wider *equivalence* (or bijection) which can be established between the two classes. Furthermore, and conversely, it turns out that the *direct* problem will in general be easier to solve, so that the embedding is in fact the other way round.

This apparent subordination of the SImpV framework to the SInsV class probably sounds counterintuitive. Its main justification is the very different kind of specification that we have used for each model. Indeed, the dynamics of the SInsV class are defined *in depth* via the chaos, while those of the SImpV framework are defined *in domain* by parametric processes. Let us now detail and contrast the two models in that respect.

Within the apparently simpler SInsV model, the value and the dynamics of the IATM volatility σ_t are *freely specifiable* using the Wiener expansion, to any required depth. So far, this is an SDE system describing adapted dynamics for only two financial instruments. To build the whole *market model* the arbitrage condition must *then* be invoked (Call prices as conditional expectations) which entirely determines the smile, both in its shape and dynamics. Unfortunately this information is, in general, not explicitly available.

By contrast, within the SImpV model the SDE specification is sparser and concerns the whole market, but it *must* be envisaged *along* its companion ZDC to carry any financial relevance, which brings a series of questions and remarks.

First, the dynamics of the SImpV coefficients ($\tilde{b}$, $\tilde{v}$ and $\vec{\tilde{n}}$) have not been provided. So a generic-depth Wiener chaos expansion of these parametric processes seems *a fortiori* out of context. Hence, is a match to the SInsV class impossible, making both the direct and inverse problems nonsensical? The answer is no, as we will see that at the IATM point the *chaos* specification of $\tilde{\Sigma}$ alone can translate into a *differential set* of all four SImpV functionals.

The second question is the degree of redundancy among these four SImpV functionals, in the perspective of *actual model parametrisation*: which are our degrees of freedom? Indeed, the ZDC is a stochastic PDE invoking multiple processes, so we must select a single dependent variable. The drift seems the best candidate, which explains why we presented the ZDC with $\tilde{b}$ on the l.h.s. The endogenous coefficient $\tilde{v}$ appears both squared and differentiated, while the exogenous coefficient $\vec{\tilde{n}}$ only exposes its modulus, so that even in a bi-dimensional framework both are more difficult to infer. Finally, the shape functional $\tilde{\Sigma}$ seems a non-starter, as having to solve dynamically a non-linear parabolic PDE is not an encouraging prospect.

The next subject is whether we have overlooked any more SImpV constraints. There are indeed some well-defined global restrictions, corresponding to the usual *validity conditions* of the smile (intra- and inter-expiry) satisfied statically but *a.s.* and at any time. We have intentionally not exploited these stochastic partial differential *inequalities*, because from an asymptotic perspective they do not bring additional *and* relevant information. They will, however, intervene in the whole-smile extrapolations covered in Chap. 4.

To link these two models and solve both the direct and the inverse problems, we must now manipulate further the main ZDC, and focus on the IATM point where the SImpV constraints are maximal. Indeed, the core principle of ACE is inductive: it involves cross-differentiating that stochastic PDE, and then taking its IATM limit under sufficient regularity assumptions.

2.3 Recovering the Instantaneous Volatility: The First Layer

After establishing some local constraints of the SImpV model, we show how to recover the associated SInsV model, then comment on and interpret this rich relationship.

2.3.1 Computing the Dynamics of σ_t

Our ultimate objective in this section is to offer a solution to the *inverse problem*, which is to *recover* the value and dynamics of the instantaneous stochastic volatility

σ_t , associated to a given SImpV model. In fact the achievements of this section will be relatively modest since, although the basics of the methods will be laid down, in effect we will only access the top level coefficients in the chaos expansion describing the dynamics of σ_t , which accounts for the mention of a “first layer”.

First we derive a collection of IATM arbitrage constraints for the SImpV class.

Proposition 2.2 (IATM arbitrage constraints of the SImpV model: first layer) *Let us consider a given sliding SImpV model, as defined by (2.11)–(2.12). Then its dynamic coefficients are locally constrained to satisfy, at the IATM point $(t, 0, 0)$:*

$$\tilde{v}(\star) = 2\sigma_t \tilde{\Sigma}'_y(\star) \quad (2.37)$$

$$\tilde{v}'_y(\star) = \tilde{\Sigma}'^2_y(\star) + \frac{3}{2}\sigma_t \tilde{\Sigma}''_{yy}(\star) - \frac{1}{2} \frac{\|\vec{n}(\star)\|^2}{\sigma_t^2} \quad (2.38)$$

$$\tilde{b}(\star) = 2\tilde{\Sigma}'_\theta(\star) - \frac{1}{2}\sigma_t^2 \tilde{\Sigma}''_{yy}(\star) + \sigma_t \tilde{v}'_y(\star) - \frac{1}{2}\sigma_t \tilde{v}(\star). \quad (2.39)$$

Proof Since we are not interested here in any θ -differential, we can take the limit of the ZDC in $\theta = 0$ first, before applying any y -differentiation. In other words, we can deal directly with the simpler IZDC instead.

Computation of $\tilde{v}(t, 0, 0)$

Let us differentiate the Immediate Zero-Drift Condition (2.28) once w.r.t. y . Omitting the arguments by assuming that all functionals are taken in $(t, y, 0)$, we obtain

$$\begin{aligned} 0 &= 4\tilde{\Sigma}^3 \tilde{\Sigma}'_y - 2\sigma_t^2 \tilde{\Sigma} \tilde{\Sigma}'_y - 2\sigma_t \tilde{\Sigma} \left[\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right] \\ &\quad - 2y \left[\sigma_t \tilde{\Sigma}'_y \left(\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right) + \sigma_t \tilde{\Sigma} \left(\tilde{v}'_y - \sigma_t \tilde{\Sigma}''_{yy} \right) + \left(\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right)^2 + \|\vec{n}\|^2 \right] \\ &\quad - 2y^2 \left[\left(\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right) \left(\tilde{v}'_y - \sigma_t \tilde{\Sigma}''_{yy} \right) + \vec{n}^\perp \cdot \vec{n}'_y \right]. \end{aligned}$$

Evaluating this at $(\star) \triangleq (t, 0, 0)$ and using the IATM identity (2.36), we get

$$0 = 4\sigma_t^2 \tilde{\Sigma}'_y(\star) - 2\sigma_t^2 \tilde{\Sigma}'_y(\star) - 2\sigma_t \left[\tilde{v}(\star) - \sigma_t \tilde{\Sigma}'_y(\star) \right],$$

which after simplification proves (2.37).

Computation of $\tilde{v}'_y(t, 0, 0)$

Differentiating the IZDC (2.28) twice w.r.t. y yields, with all functionals in $(t, y, 0)$:

$$\begin{aligned} 0 &= 12\tilde{\Sigma}^2 \tilde{\Sigma}'^2_y + 4\tilde{\Sigma}^3 \tilde{\Sigma}''_{yy} - 4\sigma_t \tilde{\Sigma} \tilde{v}'_y + 2\sigma_t^2 \tilde{\Sigma} \tilde{\Sigma}''_{yy} - 2 \left(\tilde{v}^2 + \|\vec{n}\|^2 \right) \\ &\quad - 2y \left[\sigma_t \tilde{\Sigma} \tilde{v}''_{yy} - 2\sigma_t \tilde{\Sigma}'_y \tilde{v}'_y - 3\sigma_t \tilde{\Sigma}''_{yy} \tilde{v} + \sigma_t^2 \tilde{\Sigma}'_y \tilde{\Sigma}''_{yy} - \sigma_t^2 \tilde{\Sigma} \tilde{\Sigma}'''_{yyy} \right. \\ &\quad \left. + 4 \left(\tilde{v} \tilde{v}'_y + \vec{n}^\perp \cdot \vec{n}'_y \right) \right] \end{aligned}$$

$$- 2y^2 \left[\left(\tilde{v}'_y - \sigma_t \tilde{\Sigma}''_{yy} \right)^2 + \left(\tilde{v} - \sigma_t \tilde{\Sigma}'_y \right) \left(\tilde{v}''_{yy} - \sigma_t \tilde{\Sigma}'''_{yyy} \right) + \|\vec{n}'_y\|^2 + \vec{n} \perp \vec{n}''_{yy} \right].$$

Evaluating this at $(\star) = (t, 0, 0)$, using (2.36) and (2.37), we get

$$0 = 4\sigma_t^2 \tilde{\Sigma}'_y{}^2(\star) + 6\sigma_t^3 \tilde{\Sigma}''_{yy}(\star) - 4\sigma_t^2 \tilde{v}'_y(\star) - 2\|\vec{n}(\star)\|^2,$$

which after simplification gives (2.38).

Computation of $\tilde{b}(t, 0, 0)$

Evaluating the Secondary IZDC (2.29) at the IATM point yields simply

$$\tilde{\Sigma}^3 \tilde{b}(\star) = E(\star) + F'_\theta(\star). \quad (2.40)$$

Furthermore, the definition of $F(t, y, \theta)$ (2.21) leads to the differential

$$F'_\theta(t, y, \theta) = 2\tilde{\Sigma}^3 \tilde{\Sigma}'_\theta(\circ) - \sigma_t^2 \tilde{\Sigma} \tilde{\Sigma}'_\theta(\circ) - y[\cdots] - \frac{1}{2}y^2[\cdots].$$

In particular, at the IATM point we have, using (2.36):

$$F'_\theta(\star) = 2\tilde{\Sigma}^3 \tilde{\Sigma}'_\theta(\star) - \tilde{\Sigma}^3 \tilde{\Sigma}'_\theta(\star) = \tilde{\Sigma}^3 \tilde{\Sigma}'_\theta(\star). \quad (2.41)$$

Substituting (2.41) into (2.40) we get

$$\tilde{b}(\star) = 2\tilde{\Sigma}'_\theta(\star) - \frac{1}{2}\sigma_t^2 \tilde{\Sigma}''_{yy}(\star) + \sigma_t \tilde{v}'_y(\star) - \frac{1}{2}\sigma_t \tilde{v}(\star),$$

which proves (2.39) and concludes the proof. $\square$

Proposition 2.2 does warrant further interpretation, as it illustrates the (over-) specification of the sliding stochastic implied volatility model. However, those comments will be postponed until Sect. 2.3.2 in order to give more insight into the recovery results themselves. Let us now move on to the inverse problem proper.

Theorem 2.1 (Recovery of the SInsV dynamics: the first layer) *Let us consider a given sliding SImpV model, as defined by (2.11)–(2.12). Then its local IATM specification is associated to a SInsV model, as defined by (2.8)–(2.10), whose value and dynamics are recovered as follows:*

$$d\sigma_t = a_{1,t}dt + a_{2,t}dW_t + \vec{a}_{3,t}d\vec{Z}_t, \quad (2.42)$$

with

$$a_{1,t} = 2\tilde{\Sigma}'_{\theta}(\star) + \sigma_t^2 \tilde{\Sigma}''_{yy}(\star) - \sigma_t^2 \tilde{\Sigma}'_y(\star) + \sigma_t \tilde{\Sigma}'_y{}^2(\star) - \frac{\|\vec{n}(\star)\|^2}{2\sigma_t} \quad (2.43)$$

$$a_{2,t} = 2\sigma_t \tilde{\Sigma}'_y(\star) \quad (2.44)$$

$$\vec{a}_{3,t} = \vec{n}(\star) \quad (2.45)$$

$$a_{22,t} = 2 \left[\tilde{\Sigma}'_y(\star) \tilde{v}(\star) + \sigma_t \tilde{v}'_y(\star) \right]. \quad (2.46)$$

Note that we will shortly define the SInsV coefficients invoked here as the σ_t -(2, 0) group (see Definition 2.1 [p. 57]).

Proof The fundamental IATM identity (2.36) is static but valid *a.s.* and at any time t , while the parameters y and θ are constant. Therefore it also provides us with the following dynamics:

$$\begin{aligned} d\sigma_t &= d\tilde{\Sigma}(t, y = 0, \theta = 0) = \tilde{b}(t, 0, 0)dt + \tilde{v}(t, 0, 0)dW_t \\ &\quad + \vec{n}(t, 0, 0)^\perp d\vec{Z}_t. \end{aligned} \quad (2.47)$$

By uniqueness of the decomposition and invoking Proposition 2.2 we get:

- The $a_{2,t}$ coefficient through (2.37), which provides (2.44).
- The $\vec{a}_{3,t}$ coefficient directly, which provides (2.45).

As for identifying the drift coefficient $a_{1,t}$, we wish to express it as a function of:

- The shape function $\tilde{\Sigma}$ and its differentials, all taken at the origin point $(t, 0, 0)$;
- The exogenous coefficient $\vec{n}(t, 0, 0)$.

It suffices to replace both $\tilde{v}(\star)$ and $\tilde{v}'_y(\star)$ by their respective expressions (2.37) and (2.38) in the IATM drift expression (2.39) to obtain

$$\begin{aligned} \tilde{b}(\star) &= 2\tilde{\Sigma}'_{\theta}(\star) - \frac{1}{2}\sigma_t^2 \tilde{\Sigma}''_{yy}(\star) + \sigma_t \tilde{\Sigma}'_y{}^2(\star) + \frac{3}{2}\sigma_t^2 \tilde{\Sigma}''_{yy}(\star) \\ &\quad - \frac{1}{2} \frac{\|\vec{n}(\star)\|^2}{\sigma_t} - \sigma_t^2 \tilde{\Sigma}'_y(\star), \end{aligned}$$

which after simplification provides (2.43). We have now covered all first-depth coefficients involved in the dynamics (2.42) of the instantaneous volatility σ_t . Moving on to the second depth, we then compute the dynamics of $a_{2,t}$ in order to extract the endogenous $a_{22,t}$ coefficient. From both sides of (2.37) and (2.44) it follows that

$$da_{2,t} = d\tilde{v}(\star) = [\cdot]dt + 2\tilde{\Sigma}'_y(\star)\tilde{v}(\star)dW_t + 2\sigma_t\tilde{v}'_y(\star)dW_t + [\cdot]d\vec{Z}_t$$

and we get the final result

$$a_{22,t} = 2 \left[\tilde{\Sigma}'_y(\star) \tilde{v}(\star) + \sigma_t \tilde{v}'_y(\star) \right],$$

which concludes the proof. $\square$

2.3.2 Interpretation and Comments

We shall first examine Proposition 2.2 in the perspective of SImpV model specification, before focusing on the Recovery Theorem 2.1, which answers the inverse problem. For practical reasons that shall become obvious in Chap. 3, we start by gathering the terms invoked by both results into two natural and consistent groups.

Definition 2.1 (*First Layer: the $\tilde{\Sigma}$ -(2,0) and σ_t -(2,0) groups*) The $\tilde{\Sigma}$ -(2,0) group comprises the following collection of IATM differentials:

$$\underbrace{\tilde{\Sigma}(\star) \quad \tilde{\Sigma}'_y(\star) \quad \tilde{\Sigma}''_{yy}(\star) \quad \tilde{\Sigma}'_{\theta}(\star)}_{\text{static coefficients}} \quad \underbrace{\overrightarrow{\tilde{n}}(\star) \quad \tilde{v}(\star) \quad \tilde{v}'_y(\star) \quad \tilde{b}(\star)}_{\text{dynamic coefficients}}$$

The σ_t -(2,0) group corresponds to the following coefficients of the SinsV model:

$$\sigma_t \quad a_{1,t} \quad a_{2,t} \quad \overrightarrow{a}_{3,t} \quad a_{22,t}$$

More generally, and as we shall confirm in the next section, the $\tilde{\Sigma}$ -(2,0) and σ_t -(2,0) collections constitute together what we will call the *first layer*. This terminology comes by reference to the successive *differentiation stages* of the ZDC, which are necessary to establish the asymptotic results (whether inverse or direct) as will be demonstrated in Chap. 3. In this case the suffix (2, 0) indicates that the ZDC had to be differentiated twice w.r.t. y and not at all w.r.t. θ , which is clear from the proof of Proposition 2.2.

2.3.2.1 Specification of the Stochastic Implied Volatility Model

Let us first comment *globally* on the trio of arbitrage constraints (2.37)–(2.39), before covering them individually.

First we note that they represent only *necessary* conditions of no-arbitrage. It is indeed possible to express more conditions of the same type, invoking further y - and θ -differentials of the parametric processes $\tilde{\Sigma}$, $\tilde{b}$, $\tilde{v}$ and $\overrightarrow{\tilde{n}}$, also taken at the IATM point $(t, 0, 0)$: this will become clearer in the course of Chap. 3.

Note also that we have presented the three equations in such a way that the IATM drift and endogenous volatility are expressed as functions of only the local *static*

differentials and of the exogenous coefficients $\vec{n}$. This particular presentation is obviously artificial, and mainly serves the purpose of avoiding a chaos expansion at higher orders. In reality it consists in a simple, local manifestation of the ZDC (2.18), which itself stresses the growing over-specification of the SimpV model class, as we close on the IATM point.

Indeed, the model is defined by the four functionals $\tilde{\Sigma}$, $\tilde{b}$, $\tilde{n}$ and $\vec{n}$, and the NAO condition dictates that in the full domain (t, y, θ) , one (and only one) of these functionals is redundant. Then in the Immediate sub-domain $(t, y, 0)$ the IZDC (2.28) and/or the Secondary IZDC (2.29) demonstrate that more restrictive conditions apply, involving a larger number of descriptive functionals: indeed, where θ is null the functionals $\tilde{\Sigma}(t, y, 0)$ and $\tilde{\Sigma}'_{\theta}(t, y, 0)$, for instance, are no longer redundant. Finally, at the IATM point $(t, 0, 0)$ the new Theorem 2.1 shows that the constraints become even stronger, involving a greater number of descriptors/functionals.

Therefore the presentation of Proposition 2.2 and of Theorem 2.1 is not unique, so that alternative expressions exist that will characterise a SimpV model, or at least its $\tilde{\Sigma}$ -(2,0) group of IATM differentials. Defining the model entirely through these differentials would indeed be natural, should the surface be designed using Taylor expansions for instance.¹⁴ We observe that this group contains eight processes, which are themselves constrained by three no-arbitrage conditions, so that the specification only retains five degrees of freedom. Put simply, the following IATM differentials are interchangeable by pairs:

$$\tilde{v}(\star) \quad \text{vs} \quad \tilde{\Sigma}'_y(\star) \qquad \tilde{v}'_y(\star) \quad \text{vs} \quad \tilde{\Sigma}''_{yy}(\star) \qquad \tilde{b}(\star) \quad \text{vs} \quad \tilde{\Sigma}'_{\theta}(\star)$$

Note that in all three cases, we have equivalence between a static coefficient and a dynamic one, but the endogenous IATM coefficient $\vec{n}(\star)$ is intrinsic and remains unmatched. In summary, the $\tilde{\Sigma}$ -(2,0) group can be uniquely and equivalently defined (and therefore the SimpV model characterised) by no less than eight equivalent, “minimal” configurations. Following these global considerations, let us comment individually on the three NAA constraints.

Remark 2.5 The IATM arbitrage constraint (2.37) shows that in any (sliding) stochastic implied volatility (SimpV) model, the IATM skew is HALF the Lognormal endogenous IATM coefficient.

This is certainly, once again, a strong design constraint for modellers. Equivalently, it means that if two models exhibit (possibly on purpose) the same IATM level *and* IATM skew, then the volatilities of that IATM level will necessary share the same endogenous component. This feature clearly has strong hedging implications, in particular in terms of Delta.

Let us now look at (2.38) which provides $\tilde{v}'_y(t, 0, 0)$. The latter IATM differential can be viewed in one of two ways: either as describing the variation of (endogenous)

¹⁴ In fact, as we shall discuss in Sect. 2.4, this is generally a bad idea as one wishes that surface to be initially *valid*, and also to stay so dynamically.

volatility w.r.t. strike, or as the endogenous coefficient in the dynamics of the IATM skew, which we will adopt. Indeed, assuming enough regularity¹⁵ we have

$$d\tilde{\Sigma}'_y(\star) = \tilde{b}'_y(\star)dt + \tilde{v}'_y(\star)dW_t + \vec{\tilde{n}}'_y(\star)^\perp d\vec{Z}_t. \quad (2.48)$$

In consequence, this term *will* determine a significant part of the risk for skew-based products such as ATM Binaries, Call Spreads, Risk Reversals or Collars. Furthermore, with respect to smile specification and behaviour, we observe that:

Remark 2.6 The IATM arbitrage constraint (2.38) shows that in any SImpV model, at the IATM point the endogenous Normal volatility of the skew $\tilde{v}'_y(\star)$ increases with the curvature $\tilde{\Sigma}''_{yy}(\star)$ and with the square of the skew $\tilde{\Sigma}'_y(\star)$ itself (all else equal). Note that the notion of curvature depends on the chosen variable: a smile which is convex in strike K is not necessarily so in log-moneyness y (see (B.3), p. 431).

Finally, and in a similar fashion, (2.39) shows how the drift of the IATM volatility depends on the exogenous volatility $\vec{\tilde{n}}(\star)$, but also on many static descriptors. In a way, it looks as if the IATM implied volatility is “riding the smile”. Indeed, if the slope $\tilde{\Sigma}'_\theta(\star)$ is positive, then it brings a positive component to the drift, just as if the IATM point was progressing along a static copy of the smile. In the same spirit, a convex smile (in y coordinates) will also bring a positive component: if the IATM point was riding a convex surface, then Itô’s Lemma would generate a positive drift. However, interpreting the influence of the skew and of the exogenous volatility seems a bit more arduous for now.

2.3.2.2 Recovery of the Stochastic Instantaneous Volatility Model

The IATM Identity with (2.36) and the Recovery Theorem 2.1 with (2.42) provides, respectively, the value (2.42) and the first-depth dynamics of the instantaneous volatility σ_t . Therefore they do provide together the first elements of a solution for the inverse problem, as stated initially in Sect. 2.1.4. It is also comforting to note that these results are perfectly compatible with those of [12] (see Appendix E of [35]). Also on the subject of model correspondence, this recovery brings the following significant result:

Corollary 2.3 (Injectivity from $\tilde{\Sigma}-(2,0)$ to $\sigma_t-(2,0)$) *Starting from a given SImpV model and inferring the associated SInsV model, using the notations of Definition 2.1 we have that*

The function $\tilde{\Sigma} - (2, 0) \mapsto \sigma_t - (2, 0)$ is a.s. injective.

¹⁵ More on this point in the next section.

Note that it is mapping a group of five processes into another group of five processes, that $\tilde{\Sigma}-(2,0)$ presents equivalent configurations and that the non-nullity of $\tilde{\Sigma}(\star)$ and σ_t has been assumed for convenience.

Proof Let us establish the property in a sequential fashion. Starting from any equivalent configuration of the $\tilde{\Sigma}-(2,0)$ group, $\tilde{\Sigma}(\star)$ uniquely defines σ_t through (2.36). Then either $\tilde{\Sigma}'_y(\star)$ or $\tilde{v}(\star)$ provides a_2 without ambiguity *via* respectively (2.44) or (2.37), while $\vec{a}_3$ stems directly from $\vec{n}(\star)$ *via* (2.45). The coefficient a_{22} is now uniquely defined by $\tilde{v}'_y(\star)$, or alternatively from $\tilde{\Sigma}''_{yy}(\star)$ by invoking (2.38), *via* (2.44). Finally, a_1 identifies with $\tilde{b}(\star)$ which itself is now uniquely defined by $\tilde{\Sigma}'_\theta(\star)$ as per (2.43) or (2.39).

Note that this inductive method is made possible by the *a.s.* positivity of σ_t , which is provided by the previous assumption (2.14), but this is an artificial constraint that can be lifted by considering the instantaneous variance instead of the volatility. $\square$

At this point, the natural question to raise is whether we actually face a bijection between the two groups, which would establish the beginning of an equivalence between the SinsV and SimpV model classes themselves. We shall see in the coming Sect. 2.4.2 that the bijection does exist for Layer 1. However, and as will be discussed in Chap. 3, if the direct relationship is indeed injective (irrespective of the dimension) at any order, this is *a priori* not the case for the recovery problem.

Among all the minimal combinations for the $\tilde{\Sigma}-(2,0)$ group mentioned in Sect. 2.3.2.1, there is no escaping the inclusion of the exogenous coefficient $\vec{n}(\star)$. However, for reasons that will be made clear in the sequel, it will prove especially useful to avoid the *full* specification of that exogenous coefficient, when possible. This is why we now present the following alternative expressions for the drifts and for the exogenous coefficients.

Corollary 2.4 (Alternative expressions within the Recovery Theorem 2.1) *The recovery of the SInsV drift can be expressed without the exogenous coefficient:*

$$a_{1,t} = 2\tilde{\Sigma}'_\theta(\star) - \sigma_t^2 \left[\tilde{\Sigma}'_y(\star) + \frac{1}{2}\tilde{\Sigma}''_{yy}(\star) \right] + \sigma_t \tilde{v}'_y(\star). \quad (2.49)$$

Furthermore, in the particular case of a scalar exogenous driver Z_t , where $\eta(\tilde{n})$ denotes the sign of $\tilde{n}(\star)$ and of $a_{3,t}$, we have the exogenous coefficients as

$$a_{3,t} = \eta(\tilde{n})\sqrt{2}\sigma_t \left[\tilde{\Sigma}'_y(\star)^2 + \frac{3}{2}\sigma_t \tilde{\Sigma}''_{yy}(\star) - \tilde{v}'_y(\star) \right]^{\frac{1}{2}}. \quad (2.50)$$

Proof Expression (2.49) comes trivially from (2.39), while isolating the squared modulus $\|\vec{n}(\star)\|^2$ on the left-hand side of (2.38) and then taking the square root on either side gives us (2.50). $\square$

Turning now to the SInsV dynamics as described by (2.42), we find that they provide an interesting insight. Indeed, looking first at the drift of σ_t as per (2.43), we

observe that a convex smile with a negative skew will create an increasing trend in volatility. This geometric situation happens to be extremely common, but the overall drift will depend also on the slope: a decreasing ATM volatility will counteract the previous effect and vice-versa.

As for the endogenous volatility of volatility, according to (2.44) its modulus will increase with the skew's, which also has strong trading implications as it relates a dynamic with a static option strategy.

In spite of, or rather because of the academic relevance of these results, it is important to stress the possible choices and difficulties one might encounter when confronting them in real-life markets. In particular, we have seen that it is possible to extract the SInsV coefficients from either static or dynamic descriptors of the smile: the question is naturally which is better.

In our view, the main criteria is the precision of such IATM market data. We have already mentioned (see Sect. 2.2.2) the possible difficulty of extrapolating the data in expiry, down to $\theta = 0$, which will affect the slope $\tilde{\Sigma}'_{\theta}(\star)$. Furthermore, measuring a space differential such as the skew $\tilde{\Sigma}'_y(\star)$ is subject to the liquidity of ITM and/or OTM strikes, whereas inferring dynamic coefficients, such as the endogenous vol of vol $\tilde{v}(\star)$, suffers from sampling error. In general, the latter error tends to dominate, hence it seems preferable to derive the dynamics of the SInsV model from the *shape* descriptors of the (sliding) smile. Unfortunately, and as previously mentioned, we cannot dispense with the exogenous coefficient $\vec{n}(t, 0, 0)$, which is intrinsic to all $\tilde{\Sigma}-(2,0)$ configurations.

The second issue is obviously how to deal with conflicting information from both static and dynamic sources. One has to ponder whether the discrepancy falls within an acceptable noise range, whether the modelling of the dynamics is inappropriate (e.g. dimension of the drivers) or whether the market itself might not be arbitrage-free. However, most of these questions lie in the realm of model risk and statistical arbitrage, and are therefore left to further research.

It is naturally possible to interpret and use these results further, in a market-oriented approach. In particular, should we attempt to *specify* the shape and dynamics of the smile *ex ante*, in the fashion of stochastic *implied* volatility models, the Recovery Theorem 2.1 would give us strong guidance. We choose to postpone this discussion until Sect. 2.5 [p. 77], which is dedicated to the simultaneous interpretation of both recovery and first layer results.

2.4 Generating the Implied Volatility: The First Layer

We now turn to the *direct problem*, by assuming that the input SInsV model specifies the instantaneous volatility σ_t as per (2.8)–(2.9) while the shape and dynamics of the associated smile—described by (2.12)—are unknown. We aim at providing information on the associated SImpV model, pertaining either to its shape $\tilde{\Sigma}(t, y, \theta)$ or to the SDE coefficients $\tilde{b}(t, y, \theta)$, $\tilde{v}(t, y, \theta)$ and $\vec{n}(t, y, \theta)$. As with the Recovery

Theorem 2.1, such information is focused on the IATM point $(t, 0, 0)$ but involves differentials in the y and θ directions.

Having brushed the subject of multi-dimensionality for the spot process, it seems reasonable to stick to the scalar case: therefore in the sequel of this chapter let us assume that $n_w = 1$.

2.4.1 Computing the Immediate ATM Differentials

Let us assume the dynamics of the *stochastic instantaneous volatility* model (2.8)–(2.10), although a_{21} and $\vec{a}_{23}$ will not be needed in this section.

Theorem 2.2 (Generating the first layer's IATM differentials) *Under the SInsV model defined by (2.8)–(2.10) the following local IATM differentials for the shape and dynamics of the sliding IV surface $\tilde{\Sigma}(t, y, \theta)$ can be expressed. They constitute the $\tilde{\Sigma}$ -(2, 0) group, or first layer.*

- *Local differentials of the shape process:*

$$\tilde{\Sigma}'_y(t, 0, 0) = \frac{1}{\sigma_t} \left[\frac{1}{2} a_2 \right] \quad (2.51)$$

$$\tilde{\Sigma}''_{yy}(t, 0, 0) = \frac{1}{\sigma_t^2} \left[\frac{1}{3} a_{22} \right] + \frac{1}{\sigma_t^3} \left[\frac{1}{3} \|\vec{a}_3\|^2 - \frac{1}{2} a_2^2 \right] \quad (2.52)$$

$$\tilde{\Sigma}'_\theta(t, 0, 0) = \sigma_t \left[\frac{1}{4} a_2 \right] + \left[\frac{1}{2} a_1 - \frac{1}{6} a_{22} \right] \quad (2.53)$$

$$+ \frac{1}{\sigma_t} \left[\frac{1}{8} a_2^2 + \frac{1}{12} \|\vec{a}_3\|^2 \right]. \quad (2.54)$$

- *Local differentials of the dynamics processes:*

$$\tilde{v}(t, 0, 0) = a_2 \quad (2.55)$$

$$\tilde{v}'_y(t, 0, 0) = \frac{1}{\sigma_t} \left[\frac{1}{2} a_{22} \right] + \frac{1}{\sigma_t^2} \left[-\frac{1}{2} a_2^2 \right] \quad (2.56)$$

$$\vec{\tilde{n}}(t, 0, 0) = \vec{a}_3. \quad (2.57)$$

As some readers might find it useful to express the static shape differentials w.r.t. strike K rather than log-moneyness y , we provide the two relevant expressions (skew and curvature) which come straight from applying (B.2) and (B.3):

Corollary 2.5 (IATM skew and curvature in absolute coordinates) *At the IATM point $(t, S_t, K = S_t, T = t)$, the skew and curvature can be expressed as*

$$\Sigma'_K(t, S_t, S_t, t) = \frac{a_2}{2S_t\sigma_t} \quad (2.58)$$

$$\Sigma''_{KK}(t, S_t, S_t, t) = \frac{1}{S_t^2} \left[\frac{1}{\sigma_t} \left[-\frac{a_2}{2} \right] + \frac{1}{\sigma_t^2} \left[\frac{a_{22}}{3} \right] + \frac{1}{\sigma_t^3} \left[\frac{\|\vec{d}_3\|^2}{3} - \frac{a_2^2}{2} \right] \right]. \quad (2.59)$$

It is naturally possible to accelerate the proof of Theorem 2.2 by making use of the results within the Recovery Theorem 2.1. We choose not to do so, for three reasons. The first is that we wish to bring into focus the similarities between the direct and inverse problem. Hence it makes sense to present them *in parallel* rather than sequentially. The second reason is that we will prove in the sequel (in particular, in Chap. 3) that these two problems are actually equivalent, within reasonable assumptions. Finally, the extensibility of the ACE methodology to higher orders will be established using the direct problem, and therefore the coming proof will serve as the basic sequence for further layers.

Proof (Theorem 2.2 and Corollary 2.5) As with the Recovery Theorem 2.1, **the first step** is to take the dynamics on both sides of the IATM Identity (2.36). The identification of the coefficients between (2.9) and (2.12) provides, *via* uniqueness of the martingale decomposition:

$$\tilde{b}(\star) = a_{1,t} \quad \tilde{v}(\star) = a_{2,t} \quad \vec{n}(\star) = \vec{d}_{3,t}$$

which proves (2.55) and (2.57) immediately.

The second step is to invoke the IATM arbitrage constraints of the SImpV model (the same that were necessary to prove the Recovery Theorem 2.1) which are gathered in Proposition 2.2. Then:

- combining first (2.37) with (2.55) provides the IATM skew as in (2.51);
- from (2.38) we have the IATM curvature $\tilde{\Sigma}''_{yy}(\star)$ as an explicit function of $\tilde{v}'_y(\star)$;
- from (2.39) the IATM slope $\tilde{\Sigma}'_\theta(\star)$ is an explicit function of $\tilde{\Sigma}''_{yy}(\star)$ and/or $\tilde{v}'_y(\star)$,

which leaves us facing three variables for only two equations.

The third step is to derive new relationships by actually computing some dynamics from already established results: the strategy is to first express $\tilde{v}'_y(\star)$, which will then give us $\tilde{\Sigma}''_{yy}(\star)$ and finally $\tilde{\Sigma}'_\theta(\star)$.

Let us consider the formal dynamics of $\tilde{\Sigma}(t, y, \theta)$ within the SImpV model, and assume that the regularity conditions required by Theorem 3.1.2 of [36] (p. 75) are satisfied. In other words, let us assume that we can apply the differentiation operator ∂_y on either side of (2.12) and express the corresponding flow, as was done with (2.48). Recall that we then get

$$d\tilde{\Sigma}'_y(t, y, \theta) = \tilde{b}'_y(t, y, \theta)dt + \tilde{v}'_y(t, y, \theta)dW_t + \vec{n}'_y(t, y, \theta)^\perp d\vec{Z}_t. \quad (2.60)$$

We take this expression in $(t, 0, 0)$ and on the left-hand side replace $\tilde{\Sigma}'_y(\star)$ by its expression (2.51) as a function of a_2 and σ_t . The dynamics of $a_{2,t}$ being defined by (2.10), those of the IATM skew come as

$$d\tilde{\Sigma}'_y(\star) = d\left[\frac{a_2}{2\sigma_t}\right] = [\cdots]dt + \frac{1}{2}\left[\frac{a_{22}}{\sigma_t} - \frac{a_2^2}{\sigma_t^2}\right]dW_t + [\cdots]d\vec{Z}_t.$$

Then identifying the endogenous coefficients on either side of (2.60) we obtain

$$\tilde{v}'_y(t, 0, 0) = \frac{1}{\sigma_t}\left[\frac{1}{2}a_{22}\right] + \frac{1}{\sigma_t^2}\left[-\frac{1}{2}a_2^2\right]$$

which proves (2.56). We can then isolate the curvature $\tilde{\Sigma}''_{yy}(\star)$ in (2.38) and substitute (2.51), (2.56) and (2.57) respectively in place of $\tilde{\Sigma}'_y(\star)$, $\tilde{v}'_y(\star)$ and $\vec{n}(\star)$:

$$\begin{aligned}\tilde{\Sigma}''_{yy}(\star) &= -\frac{2}{3}\frac{1}{\sigma_t}\tilde{\Sigma}'_y{}^2(\star) + \frac{1}{3}\frac{\|\vec{n}(\star)\|^2}{\sigma_t^3} + \frac{2}{3}\frac{1}{\sigma_t}\tilde{v}'_y(\star) \\ &= -\frac{2}{3}\frac{1}{\sigma_t}\left[\frac{a_2}{2\sigma_t}\right]^2 + \frac{1}{3}\frac{\|\vec{a}_3\|^2}{\sigma_t^3} + \frac{2}{3}\frac{1}{\sigma_t}\left[\frac{1}{\sigma_t}\left[\frac{1}{2}a_{22}\right] + \frac{1}{\sigma_t^2}\left[-\frac{1}{2}a_2^2\right]\right],\end{aligned}$$

which after simplification proves (2.52). Finally, isolating the slope $\tilde{\Sigma}'_\theta(\star)$ in the drift equation (2.39), before replacing $\tilde{\Sigma}'_y(\star)$, $\tilde{\Sigma}''_{yy}(\star)$, $\tilde{v}(\star)$ and $\tilde{v}'_y(\star)$ respectively by (2.51), (2.52), (2.55) and (2.56), we obtain

$$\begin{aligned}\tilde{\Sigma}'_\theta(\star) &= \frac{1}{2}\tilde{b}(\star) + \frac{1}{4}\sigma_t^2\tilde{\Sigma}''_{yy}(\star) - \frac{1}{2}\sigma_t\tilde{v}'_y(\star) + \frac{1}{4}\sigma_t\tilde{v}(\star) \\ &= \frac{1}{2}a_1 + \frac{1}{4}\sigma_t^2\left[\frac{1}{\sigma_t^2}\left[\frac{1}{3}a_{22}\right] + \frac{1}{\sigma_t^3}\left[\frac{1}{3}\|\vec{a}_3\|^2 - \frac{1}{2}a_2^2\right]\right] \\ &\quad - \frac{1}{2}\sigma_t\left[\frac{1}{\sigma_t}\left[\frac{1}{2}a_{22}\right] + \frac{1}{\sigma_t^2}\left[-\frac{1}{2}a_2^2\right]\right] + \frac{1}{4}\sigma_ta_2,\end{aligned}$$

which after simplification provides (2.53) and concludes the proof of Theorem 2.2.

In order to prove Corollary 2.5 it suffices to use the transition formulae: combining (2.51) and (B.2) proves (2.58), while applying (B.3) gives

$$\begin{aligned}\Sigma''_{KK}(t, S_t, S_t, t) &= \frac{1}{S_t^2}\left[\tilde{\Sigma}''_{yy}(\star) - \tilde{\Sigma}'_y(\star)\right] \\ &= \frac{1}{S_t^2}\left[\frac{1}{\sigma_t^2}\left[\frac{1}{3}a_{22}\right] + \frac{1}{\sigma_t^3}\left[\frac{1}{3}\|\vec{a}_3\|^2 - \frac{1}{2}a_2^2\right] + \frac{1}{\sigma_t}\left[-\frac{1}{2}a_2\right]\right],\end{aligned}$$

which matches (2.59) and concludes the overall proof. $\square$

2.4.2 Interpretation and Comments

We now translate and extrapolate the Layer-1 results, i.e. Theorem 2.2 and Corollary 2.5, into theoretically meaningful and trading pertinent results. Naturally, the latter mainly concern the influence of each instantaneous coefficient $a_{i,t}$ on the shape and dynamics of the smile, as far as IATM differentials are concerned. Although these considerations are generally model-independent, whenever pertinent we compare our results with those available in the literature (whether exact or heuristic) for the specific class of pure local volatility models.

First we go through some general aspects of those Layer-1 results, before focusing on individual IATM differentials, in particular the (static) IATM skew and IATM curvature.

2.4.2.1 General Considerations on the First Layer

Let us first mention a peculiar feature of the $\tilde{\Sigma}-(2,0)$ group of IATM differentials, as presented by Theorem 2.2. As will be discussed in Sect. 2.5.1.1 [p. 77], it has significantly beneficial properties in terms of whole-smile extrapolation.

Remark 2.7 The first layer as per Theorem 2.2 provides the IATM differential of order 2 in strike, but only 1 in maturity (and no cross-terms, which will only appear in further layers). This distinct behaviour of the space and time differentials is clearly a consequence of Itô's formula. Not surprisingly, it is also a constant of our study: with higher-order layers, cross-terms might appear but increasing the order of differentiation w.r.t. θ by one will always drop the y -order by 2. We call this feature the “ladder effect”, which will be formally established and graphically presented in Chap. 3 (see Fig. 3.1, p. 122).

In terms of model correspondence, a noticeable consequence of Theorem 2.2 is the establishment of the following result.

Corollary 2.6 (Injectivity from the $\sigma_t-(2, 0)$ to the $\tilde{\Sigma}-(2, 0)$ group) *Starting from a given SInsV model and inferring the associated SImpV model, using the notations of Definition 2.1 we have that*

The function $\sigma_t-(2, 0) \mapsto \tilde{\Sigma}-(2, 0)$ is a.s. injective.

As discussed previously in Sect. 2.3.2.1 [p. 57], this statement has to be understood within the over-specification of the $\tilde{\Sigma}-(2, 0)$ group, which effectively possesses only five degrees of freedom.

Proof To establish this result, again we proceed sequentially: starting with a given $\tilde{\Sigma}-(2,0)$ configuration, the IATM Identity (2.36) [p. 50] uniquely sets σ_t . Then (2.51) or equivalently (2.55) provides unambiguously $a_{2,t}$, while (2.57) determines $\vec{a}_{3,t}$. The coefficient $a_{22,t}$ is then uniquely defined by either (2.52) or (2.56), and finally (2.53) unambiguously sets $a_{1,t}$. $\square$

Combining Corollaries 2.3 and 2.6 directly yields the following correspondence.

Corollary 2.7 (Bijectivity of the σ_t -(2,0) and $\tilde{\Sigma}$ -(2,0) groups)

The relationship σ_t -(2, 0) $\longleftrightarrow$ $\tilde{\Sigma}$ -(2, 0) is a.s. bijective.

We consider this result to be fundamental since, if upheld at all orders of differentiation (which it is, as will be proven in Sect. 3.1) it can establish complete correspondence between the SInsV and SImpV model classes. Indeed, assuming that the smile's static and dynamic functionals are analytic along both the y and θ coordinates (i.e. that they can be entirely determined by an infinite series of IATM differentials) then their specification is equivalent to writing “in depth” the chaos expansion of a generic stochastic instantaneous volatility model. But even without the still-to-be-proven extension to higher order, the σ_t -(2,0) $\longleftrightarrow$ $\tilde{\Sigma}$ -(2,0) correspondence can and will be used by itself, no later than in Sect. 2.5.3, which is dedicated to the “intuitive” re-parametrisation of SInsV models.

Turning now to more practical matters, as announced we shall now use the local volatility (LocVol) model class as an important benchmark, due to its simplicity but also to its academic relevance. Note that one can refer to [13] for an alternative approach to the class, in a dynamic smile context presenting similarities with our framework.

The class has previously been mentioned in Sect. 2.1.2.3 [p. 29] and we will maintain the same notation for its dynamics, as per (2.5). Recall also that although its practical usage is limited (due to flawed dynamics) its explanatory and demonstrative capacity is very strong. Also it has been thoroughly investigated in the literature, which provides a good opportunity for critically assessing our results using a simple case (more complex illustrations will follow).

All these features are mainly due to the minimal Markovian dimension, and to the alternative interpretation that they offer for the IV surface, as a conditional expectation and *via* Dupire. Evidently, since our results are generic and come in closed form, we must look for the same properties in the literature. However, to our knowledge there is no published result which is simultaneously generic, exact, explicit, valid in the full domain (or failing that in a neighborhood of the IATM point) and of a direct nature (i.e. providing $\tilde{\Sigma}(\dots)$ from $f(\cdot, \cdot)$). Hence we have selected three distinct approaches for their respective strengths.

The first natural candidate is certainly Dupire, which is expressed in price terms in [37] but can be re-parametrised using (sliding) implied volatility. The equivalent formulation can be found in [28]¹⁶ (p. 6) or adapted from [38] (p. 13) or even derived without difficulty from Dupire's formula. Using our usual notation $(o) = (t, y, \theta)$ it reads as:

$$f^2(T, K) = \frac{\tilde{\Sigma}^4(o) + 2\theta \tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta}(o)}{\left[\tilde{\Sigma}(o) - y \tilde{\Sigma}'_y(o) \right]^2 - \frac{1}{4} \theta^2 \tilde{\Sigma}^4 \tilde{\Sigma}'^2_y(o) + \theta \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy}(o)}. \quad (2.61)$$

¹⁶ Note that some non-official internet versions incorporate a minor typo in the denominator.

Note that this result is verified in the whole domain (t, y, θ) , also that it falls into the inverse category (it provides the SinsV specification from the implied volatility) and finally that it is *exact*. Overall, in our context these properties make (2.61) better suited for verification purposes, rather than derivation of the direct results.

Furthermore, since our Layer-1 results are expressed at the IATM point, which is a subset of the immediate domain $\theta \equiv 0$, it becomes pertinent to bring forward another available result, provided by [30]. It states that in a local volatility model, the immediate implied volatility at any strike is the spatial *harmonic mean* of the local volatility between S_t and K . Importantly, this is a direct and exact result, which in our context reads as

$$\tilde{\Sigma}(t, y, \theta = 0) = \left[\int_0^1 f^{-1}(t, S_t e^{sy}) ds \right]^{-1}, \quad (2.62)$$

which means that it provides the pure- y immediate differentials of the implied volatility surface, in particular at the IATM point.

Finally, we compare our results to an approximation, usually referred to as Gatheral's formula, that can be found in [39] and [38]. It relies on the very definition of local variance as a conditional expectation, and expresses the (stochastic) “forward implied variance” as a stochastic integral against dS_t , the integrand of which is written as an expansion around a particular “most probable” path (MPP). The Black implied variance itself is then taken as a time integral of that expression, but after selecting order zero for the expansion (see [38], p. 30). Within our framework, the formula reads as follows:

$$\Sigma^2(t, S_t; K, T) \approx \overline{\Sigma}^2(t, S_t; K, T) = \frac{1}{T-t} \int_t^T f^2\left(s, S_t (K/S_t)^{\frac{s-t}{T-t}}\right) ds. \quad (2.63)$$

The reasons we elected to test and compare Gatheral's formula against our asymptotic results are several. Firstly it constitutes an intuitive approach: the MPP is an easy concept, and the overall expression is linked to the notion of Brownian bridge, which is widely understood. Furthermore, the formula is popular among practitioners, not least because it provides an approximation of the implied volatility in the full domain, a feature which allows us to compute all IATM differentials of Theorem 2.2.

It is worth stressing that Gatheral's formula, unlike (2.61) and (2.62), and because it stems from a low-order expansion, is *not* an exact result. It is evident that it cannot capture with a single path the *whole* local volatility function until maturity, which *does* generate the marginal distribution. For all intents and purposes, it should therefore be considered in the current section as a heuristic.

In order to derive our asymptotic Layer-1 results, the first step with any model (class) is naturally to express the cast, i.e. the corresponding SInsV coefficients:

Lemma 2.2 (Instantaneous coefficients of the local volatility model) *In a pure local volatility model, defined by (2.5), the SInsV coefficients are*

$$\sigma_t = f(t, S_t) \quad a_{2,t} = S_t f(t, S_t) f_2'(t, S_t) \quad a_{3,t} \equiv 0 \quad (2.64)$$

$$a_{1,t} = f_1'(t, S_t) + \frac{1}{2} S_t^2 f^2 f_{22}''(t, S_t) \quad (2.65)$$

$$a_{22,t} = S_t f(t, S_t) \left[(f \cdot f_2')(t, S_t) + S_t f_2'^2(t, S_t) + S_t (f \cdot f_{22}'')(t, S_t) \right]. \quad (2.66)$$

Proof For pure local volatility models, the cast immediately gives

$$\sigma_t = f(t, S_t).$$

Therefore the dynamics of σ_t become

$$d\sigma_t = \left[f_1'(t, S_t) + \frac{1}{2} f_{22}''(t, S_t) \langle dS_t \rangle \right] dt + f_2'(t, S_t) dS_t,$$

so that

$$a_{1,t} = f_1'(t, S_t) + \frac{1}{2} S_t^2 f^2 f_{22}''(t, S_t) \quad \text{and} \quad a_{2,t} = S_t f(t, S_t) f_2'(t, S_t),$$

obviously with $a_{3,t} = 0$. We then get the dynamics of $a_{2,t}$ as

$$da_2 = [\dots] dt + \left[(f \cdot f_2')(t, S_t) + S_t f_2'^2(t, S_t) + S_t (f \cdot f_{22}'')(t, S_t) \right] dS_t,$$

so that

$$a_{22} = S_t f(t, S_t) \left[(f \cdot f_2')(t, S_t) + S_t f_2'^2(t, S_t) + S_t (f \cdot f_{22}'')(t, S_t) \right]. \quad \square$$

Note that the exogenous coefficient $a_{3,t}$ is *a.s.* null; we use this opportunity to underline that, within our more general framework, we view local volatility models simply as a special instance of “purely endogenous” models. Those are characterised by a missing $\vec{Z}_t$ driver, which make of a_2 the “full” volatility of volatility. In that vein, another very simple type of endogenous model will be presented shortly (see Example 2.5, p. 108).

We can now derive the $\tilde{\Sigma}$ -(2,0) group of IATM differentials. We start with the level by combining (2.36) and (2.64):

Corollary 2.8 (IATM level in a local volatility model) *In a pure local volatility model defined by (2.5) the IATM level is simply given by*

$$\Sigma(t, S_t; K = S_t, T = t) = \tilde{\Sigma}(t, 0, 0) = f(t, S_t). \quad (2.67)$$

Quite reassuringly, it is easily seen that all above-mentioned literature approaches (respectively (2.61), (2.62) and (2.63)) agree with the asymptotic result (2.36) on the IATM volatility level. Indeed, with (2.62) it suffices to take $y = 0$, and similarly with (2.61) the (assumed) non-negativity of both $f()$ and $\tilde{\Sigma}()$ provides the desired result in $(t, 0, 0)$.

Finally, addressing Gatheral's heuristic, we start by re-expressing (2.63): defining $u \triangleq \frac{s-t}{T-t}$ and using sliding coordinates, we get the more practical formulation

$$\overline{\Sigma}^2(t, S_t; y, \theta) = \int_0^1 f^2(t + u\theta, S_t e^{uy}) du, \quad (2.68)$$

so that the same non-negativity argument brings (2.67).

2.4.2.2 Static Skew vs Endogenous Vol of Vol

Let us now examine the surprisingly simple result (2.51) which establishes a straightforward relationship between the IATM skew $\tilde{\Sigma}'_y(\star)$ (a static, implied quantity) and the endogenous “vol of vol” $a_{2,t}$ (a dynamic, instantaneous coefficient). In simple terms, we observe that:

Remark 2.8 The direct result (2.51) shows that, in any SInsV model, the IATM skew is HALF the Lognormal endogenous volatility of volatility. The latter is defined by the convention

$$\frac{d\sigma_t}{\sigma_t} = \frac{a_1}{\sigma_t} dt + \frac{a_2}{\sigma_t} dW_t + \frac{\vec{a}_3}{\sigma_t} d\vec{Z}_t.$$

In qualitative terms, it is relatively easy to build some intuition as to why skew is an *increasing* function of the endogenous vol of vol, and furthermore shares its sign. Indeed, $a_{2,t}$ can be written as the quadratic co-variation between the underlying S_t and its volatility σ_t , when both are expressed using Lognormal conventions (formally, consider $\log(S_t)$ and $\log(\sigma_t)$):

$$\left\langle \frac{dS_t}{S_t}, \frac{d\sigma_t}{\sigma_t} \right\rangle = \sigma_t \frac{a_2}{\sigma_t} dt = a_2 dt.$$

That said, let us assume for instance that this instantaneous (effective) correlation a_2 is *negative*. Then, when the underlying S_t increases, “on average” its instantaneous Lognormal volatility will decrease, and vice-versa. For a given option maturity, this phenomenon will amplify down moves and curb upmoves, compared to an actual Lognormal process, tending (again, in law) to accumulate the marginal distribution¹⁷

¹⁷ Obviously expressed under the martingale measure.

to the *left* of the money (S_t). This in turn increases the cumulative (Binary Put) value at-the-money: recalling how this cumulative is expressed as a function of the skew, we do obtain a *negative* Lognormal skew. Note that all volatilities (instantaneous and implied) must be considered under the same convention (here Lognormal) for this rough reasoning to be valid.

For a small number of specific models, including the pure local volatility class, this result has been known for some time, albeit usually expressed with differentials w.r.t. strike K . In some of these cases, the coefficient $a_{2,t}$ is referred to as the instantaneous *effective* correlation between the underlying and its volatility. This terminology owes to the fact that the co-variation we mentioned above can be generated by either correlated drivers and/or by a functional relationship, of the type $\sigma_t = f(S_t)$. We will see later in Sect. 2.5.2 how local volatility and correlation can be combined in a general class, and how to apply Theorem 2.2.

Let us now specialise the IATM skew formula (2.51) in the pure local volatility (LocVol) framework, and compare to results available in the literature. Our asymptotic results yield:

Corollary 2.9 (IATM skew in a local volatility model) *In a pure local volatility model, defined by (2.5), the IATM skew is given by*

$$\begin{aligned}\tilde{\Sigma}'_y(t, 0, 0) &= \frac{1}{2} S_t f'_2(t, S_t) \\ \Sigma'_K(t, S_t; K = S_t, T = t) &= \frac{1}{2} f'_2(t, S_t).\end{aligned}\tag{2.69}$$

Proof It suffices to combine, respectively, (2.51) and (2.58) with (2.64) to obtain the desired result. $\square$

In other words we have:

Remark 2.9 The direct result (2.51) shows that in local volatility models, the IATM skew of the Lognormal smile is HALF the IATM S -differential of the Lognormal local volatility function.

This result has been present in the literature for some time, and usually comes as a consequence of path integral approximations, of which the simplest form is the *mid-point method*. This rough proxy consists, for short-term options close to the money, in approximating the implied volatility at strike K and maturity T by the value of the *local* volatility, taken at the centre of the $[(t, S_t), (T, K)]$ segment. Note that the volatility convention is of course identical (Lognormal) in both the implied and local dynamics. Formally, we have that if

$$K = S_t + \Delta K \text{ with } \Delta K \ll 1 \quad \text{and} \quad T = t + \Delta t \text{ with } \Delta t \ll 1$$

then the approximation reads as

$$\Sigma(K, T) \approx \widehat{\Sigma}(K, T) \triangleq f\left(\frac{1}{2}(t + T), \frac{1}{2}(S_t + K)\right),$$

which happens to bring the exact IATM skew:

$$\Sigma'_K(S_t, t) \approx \widehat{\Sigma}'_K(S_t, t) = \frac{1}{2}f'_2(t, S_t).$$

The formula is naturally a very useful hedging tool, and over time has been presented in different forms and by several authors. In [40] it is brought as a heuristic, along with a couple of other “rules of thumb”; due to the pedagogic nature of the paper, the justification there is minimal. It has been made more rigorous in other, related papers (see references) but tends to rely on an intuitive idea: that the implied volatility for strike K is a path integral of the local volatility between the spot/forward S_t and K .

This is indeed the vein used by [38] to establish (2.63), albeit in a more elaborate fashion. It relies on the fact that in a Lognormal model, the MPP between (t, S_t) and (K, T) is a direct line in *log-space*, hence the integrand’s argument in (2.63). Then for a short-expiry, close-to-the-money option, at first order the MPP becomes a direct line in the initial coordinates, while at zero order it can be approximated by the midpoint. Formally, it suffices to differentiate (2.68) once w.r.t. y and then to apply at the IATM point, obtaining sequentially

$$2\overline{\Sigma}'_y(t, S_t; y, \theta) = \int_0^1 2ff'_2(t + u\theta, S_te^{u\theta}) S_t u e^{u\theta} du \quad (2.70)$$

then

$$2\overline{\Sigma}'_y(t, S_t; 0, 0) = 2S_t f(t, S_t) f'_2(t, S_t) \int_0^1 u du.$$

Finally, invoking (2.67) and (B.2) we get (2.69), as announced.

Apart from heuristics, (2.69) has also been established rigorously. For instance [41] uses a singular perturbation approach in the case where $f(s, x)$ is a function of separable variables. The resulting expansion is indeed taken around the midpoint, and differentiating once w.r.t. K provides the desired result. In the general context, we can obtain an exact formula from [30], as the skew is a pure-space differential. We simply differentiate both sides of (2.62) once w.r.t. to y to get

$$\tilde{\Sigma}'_y(t, y, \theta = 0) = - \left[\int_0^1 f^{-1}(t, S_t e^{sy}) ds \right]^{-2} \left[\int_0^1 -f^{-2} f'_2(t, S_t e^{sy}) S_t e^{sy} ds \right]. \quad (2.71)$$

Evaluating this expression at $(t, 0, 0)$, we have

$$\tilde{\Sigma}'_y(t, 0, 0) = -f^2(t, S_t) \left[-S_t f^{-2}(t, S_t) f'_2(t, S_t) \int_0^1 s ds \right] = \frac{1}{2} S_t f'_2(t, S_t).$$

Note that, in order to check against an exact result, another alternative would be to invoke (2.61), which reassuringly provides the same answer.

2.4.2.3 Static Curvature vs Exogenous Vol of Vol

Let us now examine result (2.52), providing the curvature. We notice that the exogenous coefficient $\vec{a}_{3,t}$ has a systematic and positive effect on $\tilde{\Sigma}''_{yy}(t, 0, 0)$. This feature is not surprising as it has been established for numerous model classes. In a Black model with an independent, stochastic volatility, the expression for Volga (C.2) combined with Jensen justifies that prices for ITM and OTM strikes will increase, relatively to the money: in terms of smile, this indeed translates into a curvature. More formally, following [42] it is proven in [43] that for a specific class of bi-dimensional diffusive models, with zero correlation (i.e. when a_2 and a_{22} are null), the implied volatility is symmetric and increasing with $|y = \ln(K/S_t)|$. Let us compare these results to our asymptotic approach, which gives:

Corollary 2.10 (IATM curvature in a local volatility model) *In a pure local volatility model, defined by (2.5), the IATM curvature is given by*

$$\tilde{\Sigma}''_{yy}(t, 0, 0) = \frac{1}{3} S_t f'_2(t, S_t) - \frac{1}{6} S_t^2 f^{-1} f'^2_2(t, S_t) + \frac{1}{3} S_t^2 f''_{22}(t, S_t). \quad (2.72)$$

Proof Substituting (2.64) and (2.66) into (2.52) yields (dispensing with the argument):

$$\begin{aligned} \tilde{\Sigma}''_{yy}(t, 0, 0) &= \frac{1}{3 f^2} S_t f \left[f f'_2 + S_t f'^2_2 + S_t f f''_{22} \right] - \frac{1}{2 f^3} S_t^2 f^2 f'^2_2 \\ &= \frac{1}{3} S_t f'_2(t, S_t) - \frac{1}{6} S_t^2 f^{-1} f'^2_2(t, S_t) + \frac{1}{3} S_t^2 f''_{22}(t, S_t). \quad \square \end{aligned}$$

In order to check that result, we now compute the IATM curvature using [30].

Defining $(\bullet) \triangleq (t, S_t e^{sy})$ and differentiating (2.71) w.r.t. y gets us

$$\begin{aligned} \tilde{\Sigma}_{yy}''(t, y, 0) = & -2 \left[\int_0^1 f^{-1}(\bullet) ds \right]^{-3} \left[\int_0^1 f^{-2}(\bullet) f_2'(\bullet) S_t e^{sy} s ds \right]^2 \\ & + \left[\int_0^1 f^{-1}(\bullet) ds \right]^{-2} S_t A(t, y), \end{aligned}$$

where

$$A(t, y) \triangleq \int_0^1 \left[\left[-2f^{-3} f_2'^2(\bullet) + f^{-2} f_{22}''(\bullet) \right] S_t e^{2sy} s^2 + f^{-2} f_2'(\bullet) e^{sy} s^2 \right] \cdot ds$$

Evaluating this expression at $(t, 0, 0)$, we obtain

$$\begin{aligned} \tilde{\Sigma}_{yy}''(t, 0, 0) = & 2f^3(t, S_t) S_t^2 f^{-4} f_2'^2(t, S_t) \left[\int_0^1 s ds \right]^2 \\ & + S_t f^2(t, S_t) \left[-2S_t f^{-3} f_2'^2(\star) + S_t f^{-2} f_{22}''(\star) + f^{-2} f_2'(\star) \right] \int_0^1 s^2 ds, \end{aligned}$$

hence

$$\tilde{\Sigma}_{yy}''(t, 0, 0) = \frac{1}{4} S_t^2 f^{-1} f_2'^2(t, S_t) + \frac{1}{3} \left[-2S_t^2 f^{-1} f_2'^2(\star) + S_t^2 f_{22}''(\star) + S_t f_2'(\star) \right],$$

which after simplification yields (2.72).

Alternatively, we can compare with the IATM curvature produced by Gatheral's formula (2.63). Using the following notations

$$(\circ) = (t, S_t; y, \theta) \quad \text{and} \quad (\diamond) = (t + u\theta, S_t e^{uy})$$

we differentiate (2.70) w.r.t. y and obtain:

$$\overline{\Sigma}_y'^2(\circ) + \overline{\Sigma} \overline{\Sigma}_{yy}''(\circ) = S_t \int_0^1 \left[\left(f_2'^2 + f f_{22}'' \right) (\diamond) S_t u^2 e^{2uy} + f f_2'(\diamond) u^2 e^{uy} \right] du.$$

Evaluating this expression at the IATM point $(\star) \triangleq (t, S_t; y = 0, \theta = 0)$, we obtain

$$\overline{\Sigma}_y'^2(\star) + \overline{\Sigma}_{yy}''(\star) = \left[S_t^2 f_2'^2(t, S_t) + S_t^2 f f_{22}''(t, S_t) + S_t f f_2'(t, S_t) \right] \int_0^1 u^2 du.$$

Substituting (2.67) and (2.69) and omitting argument (t, S_t) for f -differentials yields

$$\left[\frac{1}{2} S_t f_2' \right]^2 + f(t, S_t) \overline{\Sigma}_{yy}''(\star) = \frac{1}{3} \left[S_t^2 f_2'^2 + S_t^2 f f_{22}'' + S_t f f_2' \right],$$

so that finally the IATM curvature corresponding to Gatheral's formula comes as

$$\overline{\Sigma}_{yy}''(\star) = \frac{1}{3} S_t f_2'(t, S_t) + \frac{1}{12} S_t^2 f^{-1} f_2'^2(t, S_t) + \frac{1}{3} S_t^2 f_{22}''(t, S_t). \quad (2.73)$$

Comparing (2.73) and (2.72) shows a discrepancy with the second coefficient (in bold). Indeed, the heuristics of (2.63) produce a *systematic positive bias* in the computation of the IATM curvature. This discrepancy might be due to the fact that Gatheral's formula is based on a zero-order expansion: it is entirely possible that further terms might bring the necessary correction, but this is left to further research.

2.4.2.4 Remaining IATM Differentials and General Remarks

Let us first examine the expression for the IATM slope (2.53). The fact that $\tilde{\Sigma}'_\theta(\star)$ comes as an affine function of the instantaneous drift $a_{1,t}$ is a relief, as it supports the link between statics and dynamics. Beyond that property, the most obvious feature is the specific scaling of that coefficient. Indeed, it seems natural to normalise a_1, a_2 and a_3 by the initial value of the volatility, not as a Lognormal convention, but simply to distinguish magnitude from structural issues, i.e. to dissociate scale from shape. Then the IATM skew $\tilde{\Sigma}'_y(\star)$ becomes dimensionless, while the next IATM y -differential $\tilde{\Sigma}''_{yy}(\star)$ comes in σ_t^{-1} . Likewise $\tilde{\Sigma}'_y(\star)$, which is the only *dynamic* IATM strike-differential within the first layer, can be normalised as a single power of σ_t . We will see in Sect. 3.2.2 [p. 138] that this intuitive property (σ_t -homogeneity) is carried over to further *pure-strike* IATM differentials, both static and dynamic. By contrast, the IATM slope $\tilde{\Sigma}'_\theta(\star)$ is σ_t -heterogeneous, therefore more complex and difficult to interpret. Therefore we revert back to the simpler framework of local volatility models.

Corollary 2.11 (IATM slope in a local volatility model) *In a local volatility model defined by (2.5) [p. 30] the IATM slope is*

$$\tilde{\Sigma}'_\theta(t, 0, 0) = \frac{1}{2} f_1' + \frac{1}{12} S_t f^2 f_2' - \frac{1}{24} S_t^2 f f_2'^2 + \frac{1}{12} S_t^2 f^2 f_{22}'' \quad (2.74)$$

where all differentials of $f(\cdot, \cdot)$ are taken at (t, S_t) .

Proof Substituting all of Lemma 2.2 into (2.53), all f -differentials taken at (t, S_t) , we get:

$$\begin{aligned}\tilde{\Sigma}'_{\theta}(t, 0, 0) &= \frac{1}{4}S_t f^2 f'_2 + \frac{1}{2} \left[f'_1 + \frac{1}{2}S_t^2 f^2 f''_{22} \right] \\ &\quad - \frac{1}{6} \left[S_t f^2 f'_2 + S_t^2 f f'^2_2 + S_t^2 f^2 f''_{22} \right] + \frac{1}{8} \frac{1}{f} S_t^2 f^2 f'^2_2 \\ &= \frac{1}{2} f'_1 + \frac{1}{12} S_t f^2 f'_2 - \frac{1}{24} S_t^2 f f'^2_2 + \frac{1}{12} S_t^2 f^2 f''_{22}.\end{aligned}\quad (2.75)$$

□

Obviously we cannot verify the validity of this result against the exact result (2.62) as the latter is only defined in the immediate domain. This is where Dupire's formula (2.61) can be invoked, its explicit nature being overcome by its full domain of validity. Differentiating both sides once w.r.t. T , and omitting argument (t, y, θ) for all $\tilde{\Sigma}$ differentials, we obtain

$$2ff'_1(T, K) = \frac{A'_{\theta}(\circ)}{B(\circ)} - \frac{A(\circ)B'_{\theta}(\circ)}{B^2(\circ)} \quad (2.76)$$

with

$$A(\circ) = \tilde{\Sigma}^4 + 2\theta \tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta} \quad A'_{\theta}(\circ) = 6\tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta} + 6\theta \tilde{\Sigma}^2 \tilde{\Sigma}'_{\theta}{}^2 + 2\theta \tilde{\Sigma}^3 \tilde{\Sigma}''_{\theta\theta}$$

$$B(\circ) = \left[\tilde{\Sigma} - y \tilde{\Sigma}'_y \right]^2 - \frac{1}{4} \theta^2 \tilde{\Sigma}^4 \tilde{\Sigma}'_y{}^2 + \theta \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy}$$

$$\begin{aligned}B'_{\theta}(\circ) &= 2 \left[\tilde{\Sigma} - y \tilde{\Sigma}'_y \right] \left[\tilde{\Sigma}'_{\theta} - y \tilde{\Sigma}''_{y\theta} \right] - \frac{1}{2} \theta \tilde{\Sigma}^4 \tilde{\Sigma}'_y{}^2 - \theta^2 \tilde{\Sigma}^3 \tilde{\Sigma}'_{\theta} \tilde{\Sigma}'_y{}^2 \\ &\quad - \frac{1}{2} \theta^2 \tilde{\Sigma}^4 \tilde{\Sigma}'_y \tilde{\Sigma}''_{y\theta} + \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy} + 3\theta \tilde{\Sigma}^2 \tilde{\Sigma}'_{\theta} \tilde{\Sigma}''_{yy} + \theta \tilde{\Sigma}^3 \tilde{\Sigma}'''_{yy\theta}.\end{aligned}$$

Therefore evaluating (2.76) at $(T = t, K = S_t)$ we get

$$2ff'_1(t, S_t) = 4\tilde{\Sigma} \tilde{\Sigma}'_{\theta}(\star) - \tilde{\Sigma}^3 \tilde{\Sigma}''_{yy}(\star).$$

Isolating the IATM slope and then substituting both (2.67) and (2.72) we obtain

$$\begin{aligned}\tilde{\Sigma}'_{\theta}(\star) &= \frac{1}{2} f'_1(t, S_t) \\ &\quad + \frac{1}{4} f^2(t, S_t) \left[\frac{1}{3} S_t f'_2(t, S_t) - \frac{1}{6} S_t^2 f^{-1} f'^2_2(t, S_t) + \frac{1}{3} S_t^2 f''_{22}(t, S_t) \right],\end{aligned}$$

which after simplification matches the desired result (2.74).

Note the similarity of the first term ($\frac{1}{2}f_1'$) with the IATM skew ($\frac{1}{2}f_2'$) and the mid-point method. Again it seems interesting to compare this exact result against Gatheral's heuristic formula. Differentiating (2.68) once w.r.t. θ and again defining $(\diamond) = (t + u\theta, S_t e^{u\theta})$ we obtain

$$\overline{\Sigma}'_{\theta}(t, S_t; y, \theta) = \int_0^1 f(\diamond) f_1'(\diamond) u du.$$

Then evaluating this expression at $(y = 0, \theta = 0)$ yields

$$f(t, S_t) \overline{\Sigma}'_{\theta}(t, S_t; 0, 0) = f f_1'(t, S_t) \int_0^1 u du \quad \text{so} \quad \overline{\Sigma}'_{\theta}(t, S_t; 0, 0) = \frac{1}{2} f_1'(t, S_t).$$

This expression clearly misses several terms, and ignores completely the role of the space coordinate in generating the IATM slope. However, we mitigate this structural shortcoming by the experience that, in practice, the magnitude of the ignored terms tends to be relatively small.

Let us now turn to the endogenous dynamics of the IATM skew, as described by (2.56). They show that $\tilde{v}'_y(\star)$ is totally independent from the exogenous specification (i.e. from $\vec{a}_{3,t}$ or $\vec{n}(\star)$) which was not obvious when considering the recovery result (2.38) [p. 54]. Note also that $\tilde{v}'_y(\star)$ turns out to be a *dimensionless* quantity: it will remain unchanged if the volatility level σ_t is scaled by a constant λ . However, scaling the (endogenous) vol of vol $a_{2,t}$ will have a quadratic and negative effect on $\tilde{v}'_y(\star)$.

Note that in a LV model, if a proxy formula matches the IATM skew *as a process* (i.e. dynamically) then the latter's endogenous volatility will also agree, hence $\tilde{v}'_y(t, 0, 0)$ will match.

Now taking a step back and looking globally at (2.69)–(2.72) a very natural pattern is seen to emerge: indeed, these pure-strike differentials seem to be controlled sequentially by the successive space differentials of the local volatility function f . In order to put this impression into perspective, we bring forward the following remark.

Remark 2.10 The bijectivity established by Theorems 2.1 and 2.2 between the $\tilde{\Sigma}^-(2,0)$ and $\sigma_t^-(2,0)$ groups can be modified for endogenous models. This is due to the exclusive relationship between $\vec{a}_{3,t}$ and $\vec{n}(\star)$, characteristic of the constraints imposed on the SImpV model (see Sect. 2.3.2.1 [p. 57]). Specifically, we underline the following subset of the main bijection:

$$\left[\sigma_t, a_{1,t}, a_{2,t}, a_{22,t} \right] \xrightleftharpoons[\text{Endog. Models}]{\vec{a}_{3,t} \equiv 0} \left[\tilde{\Sigma}(\star), \tilde{\Sigma}'_y(\star), \tilde{\Sigma}''_{yy}(\star), \tilde{\Sigma}'_{\theta}(\star) \right]$$

where we intentionally use only shape differentials on the right-hand side.

In particular, within LV models Corollaries 2.67, 2.69, 2.72 and 2.74 show that the IATM level, skew and curvature of the smile are sequentially and respectively controlled by the IATM space differentials of order 0, 1, 2 of the LV function.

By sequentially we mean that matching $f^{(i)}(t, S_t)$ is conditional on all $f^{(j)}(t, S_t)$ having been set, for $0 \leq j < i$. Likewise, the IATM slope is then controlled by the first time differential of the LV function. Although we haven't yet established the results to support this statement, in accordance with intuition this property is carried over to higher orders. In other words, with LV models the control over the IATM $\tilde{\Sigma}$ -differentials is established by the corresponding differentials of f .

2.5 Illustrations and Applications

In the course of solving the inverse and direct problem, our efforts so far have focused on establishing the raw asymptotic results, and providing a mathematical interpretation of the structural relationships between three groups of processes:

- The coefficients (in the chaos decomposition) of the stochastic instantaneous volatility model: $\sigma_t, a_{2,t}, a_{3,t}$, etc.
- The smile (IATM) shape descriptors: $\tilde{\Sigma}'_y(\star), \tilde{\Sigma}''_{yy}(\star)$, etc.
- The smile (IATM) dynamics descriptors: $\tilde{v}(\star), \tilde{n}(\star), \tilde{v}'_y(\star)$, etc.

We are coming now to a more applicative phase, where our previous results shall be examined with realistic modelling and trading concerns in mind. So far we have been considering SInsV and SImpV models “in parallel”, not giving precedence to one over the other: simply because the mathematical framework exhibits such a symmetry between the two classes. But at the time of writing, stochastic *implied* volatility models are rarely used in practice. Therefore, the modelling concern will tend to be focused on the *instantaneous* model class.

The results we will use and interpret are those of Theorems 2.1 [p. 55] and 2.2 [p. 62]. We will group these Theorems, their corollaries and the exposed quantities under the overall denomination of “First Layer”. Throughout this section, we make a deliberate effort to build some intuition, and also to continue introducing the topics to be developed in further chapters.

2.5.1 An Overview of Possible Applications

2.5.1.1 General Considerations

First of all, Remark 2.7 [p. 65] deserves additional comments, from the perspective of practical implementations. We observed that the IATM differentials provided by the first layer are of the second order in space (strike) but only of the first order in time

(to maturity). This is naturally a consequence of Itô's Lemma, and this discrepancy will present itself at every order of differentiation, as shown later by the Ladder Effect (in Sect. 3.1). But this is also lucky for practitioners willing to extrapolate the smile shape. Indeed, most live market smiles do exhibit a lot more variation and irregularity in the strike dimension than in the expiry one: for instance curvature is a lot more pronounced,¹⁸ so that we need $\tilde{\Sigma}_{yy}''(\star)$ more than $\tilde{\Sigma}_{\theta\theta}''(\star)$.

Also, one might wonder about the consequences of extrapolating the implied volatility far from the money, using Maclaurin series for instance. Again, in practice we are helped by the fact that Vega,¹⁹ the sensitivity of price w.r.t. Lognormal implied volatility, vanishes when the strike goes to zero or to infinity. This might not have been the case, or maybe not in such an obvious manner, had we used another implied parameter than Lognormal volatility. Beyond this mathematical argument let us recall that, for these far-from-the-money strikes, liquidity drops and therefore bid/ask spreads increase (at least in relative terms). This market feature does mitigate the fact that the attainable precision is comparatively lower in these regions.²⁰

An attractive feature of this asymptotic approach is its generic nature: the fact that our results are produced for abstract stochastic volatility models, both instantaneous and implied. Also, as will be proven in Chap. 3, the methodology can be extended to any differential order.

Note also that with a single underlying and a scalar endogenous driver, we can directly apply our asymptotic results to the most common class of *mixtures* models, those combining *prices* with *fixed weights*. However, as a model class as well as a technical tool, mixtures constitute a wider family and do offer interesting (but not necessary satisfactory) properties: see [44, 45] and [46], among others. They are also strongly linked to the notion of a basket, which we will cover in Chap. 3, both in the constant and in the stochastic weights configurations (see Sect. 3.5 [p. 185]).

Despite these positive aspects, the asymptotic nature of the results brings structural limitations. Indeed, the model is “localised” at the IATM point. If, for instance, we are dealing with a SinsV model of the parametric diffusive class (the most common type) then all functional coefficients of the SDE will be seen through their partial differentials, taken at the initial point. True, if these functions are analytic, then pushing our method to an infinite order would theoretically solve the issue. Obviously in practice this is not an option, so that large variations of these maps w.r.t. time (non-stationarity) and/or space will degrade the output quality, at a given order of differentiation: ideally, we would like an integral approach as in [47].

There is, however, in the modelling community, a strong argument for stationary or time-homogeneous models (see [48] for instance). We support that view, within

¹⁸ One could argue that this is a natural consequence of the higher liquidity of strike-based products (Butterflies, Strangles, etc.) compared to maturity spread products, which itself leads to the question of forward volatilities, but this is out of our current scope.

¹⁹ $\mathcal{V} = S_{t,T} \mathcal{N}'(d_1)$.

²⁰ Note that the notions of precision and sensitivities must be clearly defined. In particular one might choose to focus either on absolute or on relative precision.

reason and while maintaining some flexibility,²¹ because in practice stationary models tend to stabilise the calibration process and the hedge. Indeed, not relying on time-dependency to improve the calibration forces the modeller to develop deeper, more involved and hopefully more realistic dynamics. *A contrario*, heavily using time functions to fudge a good fit means that once the market has moved, the new calibrated time structure is likely to be very different. Although the later approach is easy to implement and will effectively provide an efficient calibration of the benchmark products, it will also create wild swings for the parameters and for most other products that are not included in the calibration set.

To some extent, the presence of *mean-reversion* will also require to obtain higher-order differentials, as it distorts the relationship between instantaneous and averaged dynamics. Another clear limitation of the method is that it does not envisage jumps: this is left for further research, as it requires a very different technical context.

Overall, when using these results we usually end up in one of two situations:

- either we exploit the raw, asymptotic equations linking IATM differentials and instantaneous coefficients; or
- we exploit these equations within *extrapolation* schemes, usually providing the whole smile, which tends to involve a large proportion of “engineering” skills.

Let us briefly detail these two approaches, before providing some examples.

2.5.1.2 Pure Asymptotic Applications: Qualitative Approaches

Let us first turn to asymptotic applications, starting with what we shall call *qualitative model design & analysis* and which clearly refers to the direct problem.

Most practitioners rely on stochastic instantaneous volatility models for pricing and hedging. It is therefore in their interest to better understand these models’ behaviour, in particular the influence or cross-interferences of the various parameters (vol of vol, correlation, mean-reversion, etc.) or of specific functional forms (local volatility, time-dependent parameters, etc.) on the volatility surface, on the joint dynamics of the underlying with its instantaneous volatility, or with the smile, etc.

Such a precise understanding enables the agent to deliver a better hedge, and the modeller to customise an existing model or even design a new one *ex nihilo* in order to fulfil given trading needs. However, most currently used SInsV models²² depend on numerical engines to price, whether it be Finite Differences or Monte-Carlo schemes, or even (Fast) Fourier Transform. Hence the difficulty with judging the impact of modelling choices on the smile, both in static and dynamic terms.

Alternatively, the approach that we advocate consists in focusing on the IATM region and using a low-level differential approach, by manipulating the usual and

²¹ In particular in the absence of economic rationale, which could distinguish between time periods.

²² With the noticeable exception of SABR.

meaningful smile descriptors which are level, skew, curvature, slope, etc. Clearly the $\tilde{\Sigma}-(2,0)$ group generated by Theorem 2.2 provides most of these. We will show in Chap. 3 that, providing simple and realistic assumptions, all IATM differentials can be expressed. In particular, Sect. 3.2 will give the most important ones, such as the twist and flattening (see the introduction of that chapter for a *typology* of the smile). Very shortly, Sect. 2.5.2 will be dedicated to an illustration of that approach. It will focus on the use of skew functions within stochastic volatility models, and in particular on the comparison of the Lognormal Displaced Diffusion and CEV instances.

Still in their pure asymptotic form, we can also use the formulas in an *inverse* manner. We can either exploit the recovery formulas provided by the Recovery Theorem 2.1, or try and invert the direct expressions of Theorem 2.2. The point is usually to re-parametrise, either totally or partially, a stochastic instantaneous volatility model by using meaningful, market-related quantities associated to the smile. As for model analysis, the rationale is the difficulty of appreciating the magnitude or impact of parameters in typical SinsV models, such as mean-reversion or vol of vol. Therefore the inverse method transforms the original model into an “intuitive” version, which is parametrised *via* its most pertinent *output*, the smile. Sect. 2.5.3 will illustrate this method on several simple examples.

2.5.1.3 Whole Smile Extrapolations and the Pertinence of Polynomials

When considering the direct problem, there is a great temptation to extrapolate the IATM differentials that have been computed (in particular those given by Theorem 2.2) to the whole smile, i.e. for all strikes and all maturities.

The natural method to exploit these results is to develop Taylor/Maclaurin series with both strike and maturity as variables. This is indeed the approach taken by [12] to describe the absolute surface $\Sigma(t, S_t, K, T)$, and we propose now to mimic that approach.

Within our framework, we denote by $\tilde{\Sigma}^*(t, y, \theta)$ the polynomial approximation of the sliding implied volatility $\tilde{\Sigma}(t, y, \theta)$, matched at the IATM point $(t, 0, 0)$. Let us invoke the IATM Identity (2.36) along with the static results of Theorem 2.2: (2.51), (2.52) and (2.53). We then write the Maclaurin series as

$$\begin{aligned} \tilde{\Sigma}^*(t, y, \theta) = & \sigma_t + y \frac{a_2}{2\sigma_t} + \frac{1}{2} y^2 \frac{1}{\sigma_t^3} \left[\frac{1}{3} \sigma_t a_{22} - \frac{1}{2} a_2^2 + \frac{1}{3} \|\vec{a}_3\|^2 \right] \\ & + \theta \left[\frac{1}{2} a_1 + \frac{1}{4} \sigma_t a_2 + \frac{1}{8} \frac{a_2^2}{\sigma_t} + \frac{1}{12} \frac{\|\vec{a}_3\|^2}{\sigma_t} - \frac{1}{6} a_{22} \right]. \end{aligned} \quad (2.77)$$

As for approximating the static *absolute* implied volatility surface, we can use Corollary 2.5, namely (2.36), (2.58), (2.59) and (2.53) to obtain

$$\begin{aligned}
\Sigma^*(t, S_t, K, T) = & \sigma_t + \left[\frac{K}{S_t} - 1 \right] \frac{a_2}{2\sigma_t} \\
& + (T - t) \left[\frac{a_1}{2} + \frac{a_2}{4}\sigma_t + \frac{1}{8}\frac{a_2^2}{\sigma_t} + \frac{1}{12}\frac{\|\vec{d}_3\|^2}{\sigma_t} - \frac{a_{22}}{6} \right] \\
& + \frac{1}{2} \left[\frac{K}{S_t} - 1 \right]^2 \left[\frac{1}{\sigma_t^3} \left[\frac{1}{3}\sigma_t a_{22} - \frac{1}{2}a_2^2 + \frac{1}{3}\|\vec{d}_3\|^2 \right] - \frac{a_2}{2\sigma_t} \right].
\end{aligned} \tag{2.78}$$

Quite reassuringly, these results are totally compatible with [12], as proven in Appendix E of [35]. It is important to note, however, that this expression cannot be used confidently for trading the vanillas, not necessarily because it constitutes only a low-level approximation,²³ but because it cannot *a priori* be guaranteed as valid, even statically. Besides, this is only part of the story, as the dynamic coefficients $\tilde{b}$, $\tilde{v}$ and $\vec{\tilde{n}}$ also need to ensure the validity of the surface in the future, as well as satisfy the ZDC everywhere. Finally, as *approximations of the dynamics*, they must tally the *dynamics of the (static) approximations*: in a nutshell, both groups have to be *consistent*.

At this stage of the study, we choose not to dwell on the various flavours and difficulties of these extrapolations. Some basic considerations will be discussed later in Sect. 4.1 and a practical application will be presented in Sect. 4.5. For now, it suffices to say that Taylor/Maclaurin series on the implied volatility and w.r.t. strike are certainly not the only option, and unfortunately are not as straightforward as we might hope. Also, we can already anticipate that the two hurdles that any approximation will have to tackle are, on one hand, the *validity* of the associated price surface (both statically and dynamically), and, on the other hand, the *precision* of that proxy w.r.t. the real model.

The fact that we are dealing with such approximations argues in itself against pure pricing applications: the risk of inconsistency and of arbitrage could indeed be significant. But on the other hand, these extrapolations can provide good initial guesses that can be used in static or dynamic calibration procedures: indeed, most *global* calibration procedures use an optimiser, which itself heavily invokes the actual pricer. More precisely, having selected a (large) collection of options $[K, T]$ for which market/target prices are available, then for a given set of model parameters the *market error* is defined as a metric²⁴ between the model option prices and the corresponding targets. The optimisation engine is tasked with minimising the market error, as a function of the model parameters, which are our (usually constrained) variables. Any such engine will therefore make numerous calls to the pricer, which itself for complex models usually consists in another numerical engine: FFT, PDE solver (finite differences or finite elements) or Monte-Carlo. That pricer will therefore be much slower than an extrapolation formula based on our asymptotic results, coming in closed form.

²³ Higher-orders differentials should be required when far from the money.

²⁴ Typically weighted squared differences, but some exact bootstrap is often involved.

However, the speed, the stability and possibly the convergence of the method will depend on a good starting point.²⁵ Therefore, by first running the optimisation engine on the approximate pricer we can attain a very good starting point, in a fraction of the time required by the nominal procedure. Also, before handing over to the real optimiser, the approximation formulas can provide good initial approximations of the Jacobian or of the Hessian. Indeed, this feature becomes very useful when the optimiser/root-finder relies on Gauss-Newton or Quasi-Newton algorithms (typically BFGS, Broyden, or Levenberg-Marquardt: see [49]). The proxy also enables us to address the possible instability issue, i.e. the existence of local minima: its speed enables it to be incorporated in dedicated methods, such as genetic algorithms.

In terms of fast calibration, note that the pure asymptotic results can also be used, as will be discussed in Sect. 2.5.3. Less precise, but faster than whole smile extrapolations, these asymptotic methods can therefore be employed to initialise the procedure described above, thus solving for the optimal parameters “in cascade”.

2.5.1.4 Sensitivities and Hedge Ratios: Delta, Gamma, Vega & Co.

In both fields of application mentioned above, i.e. pure asymptotics and extrapolations, the IATM differential results in general, and the first layer in particular, grant access to more than just the smile’s shape and dynamics. Indeed, they also provide hedging and risk information in the form of many *hedge ratios*, including Delta, Gamma, Vega, Volga and Vanna (see p. 44 for definitions).

As will shortly be discussed, in a market completed with options it is ultimately the *hedge strategy* that is the real determinant in terms of pricing and overall risk management. And in the context of unobservable and/or non-tradeable state variables, in particular with SV models (both instantaneous and implied), the very notion of hedge and its relationship to various price differentials becomes ambiguous.

Hedging with parametric diffusions

With these features in mind, we start by comparing the SImpV class to the dominant type of SInsV models, i.e. *parametric diffusions*. The ultimate purpose of this model class is to describe the dynamics of the (single) underlying S_t and *a priori* of no other financial quantity, especially not of options on S_t . To express those dynamics, the model is equipped with a finite number M of auxiliary *state variables* (typically some stochastic volatility) gathered under the notation $\vec{X}_t$. Restricting ourselves to Itô

²⁵ This is particularly the case for “scratch” calibration, where the parameters are completely unknown, but less so for re-calibration, which usually starts from the previous parameter set.

processes, this dynamic system is then driven by the following *Markovian* parametric diffusion:

$$dS_t = \vec{f}(t, S_t, \vec{X}_t; \vec{\beta})^\perp d\vec{B}_t \quad (2.79)$$

$$d\vec{X}_t = \vec{g}(t, S_t, \vec{X}_t; \vec{\beta}) dt + \vec{h}(t, S_t, \vec{X}_t; \vec{\beta}) d\vec{B}_t. \quad (2.80)$$

By convention, $\vec{B}_t$ is a standard Wiener process of finite dimension $N \leq M + 1$, with unit correlation matrix (i.e. with independent components) and which aggregates both the endogenous and exogenous drivers. The measure is chosen by convention as martingale for the underlying and for all European options considered. The individual components $X_{i,t}$, on the other hand, are not necessarily driftless and can typically be mean-reverting. Finally, by $\vec{\beta}$ we denote the finite set of parameters specifying the functional coefficients $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ which define the *parametric* diffusion.

The way in which the underlying S_t as well as its dynamics (2.79)–(2.80) are generated categorises the type of model considered, which usually come in one of two flavours:

- In *Factor Models*, a finite set of processes (often the components of $\vec{X}_t$ itself) represent abstract quantities, which with time t represent the actual state variables of the dynamic system. Consequently the underlying S_t is initially absent from the defining SDE system, which typically misses (2.80). The abstract factors are then mapped onto the underlying *via* a function. To fall back into our framework, this mapping function must be parametric²⁶ as with some Linear or Quadratic Gaussian models. We will not dwell on this class, which is sometimes also referred to as *Markov Functional*.
- In *Market Models*, S_t is a *market* instrument and its dynamics (2.79) are explicitly included. Formally, if $\vec{X}_t$ is empty then we have a local volatility model (*à la* Dupire) whereas in the case where $\vec{X}_t \equiv \sigma_t$ we may have a stochastic volatility model in the usual sense, such as with Heston or SABR.

Having defined the underlying's dynamics, we now turn to the price system of European options. The latter being uniquely determined by no-arbitrage assumptions under the chosen measure, European prices²⁷ will therefore come as deterministic (but not necessarily explicit) functions of:

- The Markovian state variables t , S_t and $\vec{X}_t$ (e.g. $[t, S_t, \sigma_t]$).
- The parameter set $\vec{\beta}$ specifying the diffusion (e.g. correlation, mean-reversion, vol of vol).
- The option parameters, usually strike K and expiry T .

²⁶ General Factor models strictly contain this limited setup, e.g. some use a non-parametric mapping.

²⁷ And therefore the associated marginal distributions.

The role of the auxiliary state variables $\vec{X}_t$ is partly to bring additional drivers to the smile. This is unlike pure local volatility (LV) models, for instance, which are strictly contained within *purely endogenous (PE) models*. In the latter, *all* components of $\vec{B}_t$ affect dS_t , which allow the shape and dynamics of the smile to be functions of the $X_{i,t}$ components. Since those processes are *a priori* non-observable and therefore non-tradeable, PE models imply *a priori* an incomplete market, as opposed to their LV subset.

The parametric diffusion family can of course be cast within our SInsV framework, as specified in Sect. 2.1.3.1 [p. 35]. But, although this class is noticeably simpler than the SImpV framework, it already exhibits ambivalent definitions for Greeks and hedges. Indeed, those can come either as partial differentials of the price w.r.t. state variables and parameters, or as individual weights in some replicating portfolio. Furthermore, although that portfolio value process is unique here, as an Itô integral it can be alternatively defined against the Brownian drivers, the state variables or some vanilla options. In the last two cases the weights are generally *not* unique, which is the downside of “hedging the drivers” through proxies.

Within the most generic framework of stochastic *implied* volatility models, the picture now gets even more complex as the notion of Greek becomes significantly ill-defined. Indeed, the Call price and/or the implied volatility processes are not necessarily deterministic functions of certain state processes (including t and S_t): recall that in the general case we simply dispense with state variables, apart from t and S_t (see Sect. 2.1.3.2). Accordingly, it is not *a priori* possible to identify a single such state variable that could pass as a “volatility” and be used to define Vega, Volga and Vanna. Also, differentiating price or volatility w.r.t. S_t becomes nonsensical in the general case: indeed we are now facing stochastic functionals of S_t , all driven by a multi-dimensional Wiener process which we know to be *a.s.* nowhere differentiable.

In summary, the differential definitions of Delta, Gamma and Vanna disappear, while the concept of Vega or Volga becomes obscure. As for the definition of those Greeks as replication weights against anything else than the drivers themselves, again there is no uniqueness. This is always the case when opting for a finite set of options as hedging instruments, but made even more pregnant as we are now dealing with a continuum of these. Whereas writing the process as a stochastic integral against some state variables is now only possible for S_t , which defines an extended Delta. Let us develop this last point, since Delta is by far the most important of the Greeks. Recall then that in the SImpV context, we have the dynamics of the Call from (2.22) [p. 44] as

$$\begin{aligned} dC(t, S_t, T, K) = & \left[\Delta S_t \sigma_t + \mathcal{V} v \sqrt{\theta} \right] dW_t + \mathcal{V} \sqrt{\theta} \vec{n}^\perp d\vec{Z}_t \\ & + \left[-\mathcal{V} \Sigma (2\sqrt{\theta})^{-1} + \mathcal{V} b \sqrt{\theta} + \frac{1}{2} \Gamma S_t^2 \sigma_t^2 \right. \\ & \left. + \frac{1}{2} \vartheta \left(v^2 + \|\vec{n}\|^2 \right) \theta + \Lambda S_t \sigma_t v \sqrt{\theta} \right] dt. \end{aligned}$$

The endogenous coefficient within brackets can be seen as a form of Delta because, in principle, it is only hedgeable with S_t :

$$\left[\Delta S_t \sigma_t + \mathcal{V} \sqrt{\theta} \right] dW_t = \underbrace{S_t^{-1} \sigma_t^{-1} \left[\Delta S_t \sigma_t + \mathcal{V} \sqrt{\theta} \right]}_{\text{Extended Delta}} dS_t.$$

However this extended Delta involves the Vega of Black's formula, which can be confusing. Also, it is *a priori* not measurable w.r.t. the filtration generated by the underlying, or even by the endogenous driver itself. Note, however, that this was already the case with parametric diffusion models.

We conclude that w.r.t. hedging and Greeks issues, our safest option is to limit ourselves to the parametric diffusion models (thankfully representing the overwhelming majority of practical implementations) and to define Greeks most simply as differentials w.r.t. the state variables t , S_t and $\vec{X}_t$ (the *hedges*) and w.r.t. the parameters $\vec{\beta}$, K and T (the *sensitivities*). We still have to link those Greeks to our IATM asymptotic results written on the sliding implied volatility. Although in principle we have access to the whole smile through extrapolations, we will focus on exact IATM results. In consequence, it is intuitive to anticipate that the limitations in scope of the first layer (or rather, the $\tilde{\Sigma}$ -(2,0) group: order 1 in maturity and 2 in strike) will necessarily have consequences.

Greeks vs the sliding smile

The most liquid options are usually those defined with fixed strike K and fixed expiry T (the *absolute* coordinates) and their hedging will (one way or another) invoke their price sensitivities. In our asymptotic context, some of the natural questions to address are therefore:

- Whether and how the Greeks (Delta, Gamma, Vega, etc.), defined here simply as price differentials, are affected by moving to the sliding coordinates y and θ ;
- Which and how IATM *absolute* Greeks can be obtained, given the first layer results, i.e. with a specific set of IATM differentials for the *sliding* IV $\tilde{\Sigma}$.

Our roadmap is the following: the first step will be to relate *absolute price* differentials to their *sliding implied volatility* counterparts. Then in a second phase we will identify *which differentials are affected* in the general domain. Finally, our third step will focus on the *IATM option* and derive the main absolute Greeks.

Remaining within the parametric diffusion framework (2.79)–(2.80) we have both the price and the IV as deterministic (but not necessarily explicit) functions $\underline{C}$ and $\underline{\Sigma}$ of the state variables, of the model parameters and of the option:

$$\underline{C}(t, S_t, \vec{X}_t; \vec{\beta}; K, T) = B(S_t, K, \sqrt{T-t} \underline{\Sigma}(t, S_t, \vec{X}_t; \vec{\beta}; K, T)).$$

A look at Appendix C confirms that all relevant differentials of the Call price w.r.t. its arguments can be obtained (without singularity issues) by combining the differentials of Black's formula and those of $\underline{\Sigma}$. We have hence answered the first question, and we also note that the same approach could be used with a *Normal* Implied Volatility and Bachelier's formula.

We now turn to the second question of how to extract the differentials of the absolute implied volatility $\underline{\Sigma}$ from those of its sliding version $\tilde{\Sigma}$, defined and denoted as follows:

$$\underline{\Sigma}(\square) \triangleq \tilde{\Sigma}(\circ) \quad \text{with} \quad (\square) \triangleq \left(t, S_t, \vec{X}_t; \vec{\beta}; K, T\right) \quad \text{and} \quad (\circ) \triangleq \left(t, S_t; \vec{X}_t; \vec{\beta}; y, \theta\right). \quad (2.81)$$

As discussed in Sect. 2.1.2.3 [p. 29], the move to sliding coordinates has *embedded* the two state variables t and S_t into $\tilde{\Sigma}$, now a *parametric* process. Clearly (2.81) shows that the differentials w.r.t. the auxiliary state variable $\vec{X}_t$ and w.r.t. the model parameter $\vec{\beta}$ are unaffected by that slide. In particular, within a classical SInsV model such as Heston, where the volatility process σ_t belongs to $\vec{X}_t$, the Vega and Volga as well as the sensitivity to correlation and vovol are unchanged. We are therefore left with four variables which, from Black's formula and from the definition of y and θ , can be grouped by symmetry into a *time* pair (time t and maturity T) and a *space* pair (underlying S_t and strike K). We obtain the relevant Greeks as:

- **Theta** $\underline{\Sigma}'_t(\square) = \tilde{\Sigma}'_t(\circ) - \tilde{\Sigma}'_\theta(\circ)$ (T -differentials are similarly obtained).
- **Delta** $\underline{\Sigma}'_{S_t}(\square) = \tilde{\Sigma}'_{S_t}(\circ) + \tilde{\Sigma}'_y(\circ) \frac{\partial y}{\partial S_t}$.
- **Gamma** $\underline{\Sigma}''_{S_t S_t}(\square) = \tilde{\Sigma}''_{S_t S_t}(\circ) + 2\tilde{\Sigma}''_{y S_t}(\circ) \frac{\partial y}{\partial S_t} + \tilde{\Sigma}''_{yy}(\circ) \left[\frac{\partial y}{\partial S_t}\right]^2 + \tilde{\Sigma}'_y(\circ) \frac{\partial^2 y}{\partial S_t^2}$.

The K -differentials would lead to similar equations, which are valid in the whole (K, T) domain. As announced, our third step is now to focus on the IATM point, where we shall use the abbreviations

$$(\diamond) \triangleq \left(t, S_t, \vec{X}_t; \vec{\beta}; S_t, t\right) \quad \text{and} \quad (\star) \triangleq \left(t, S_t; \vec{X}_t; \vec{\beta}; 0, 0\right).$$

Evaluating the three differential expressions above at the IATM point, we get respectively

$$\begin{aligned} \text{IATM Theta} \quad & \underline{\Sigma}'_t(\diamond) = \tilde{\Sigma}'_t(\star) - \tilde{\Sigma}'_\theta(\star). \\ \text{IATM Delta} \quad & \underline{\Sigma}'_{S_t}(\diamond) = \tilde{\Sigma}'_{S_t}(\star) - \frac{1}{S_t} \tilde{\Sigma}'_y(\star). \\ \text{IATM Gamma} \quad & \underline{\Sigma}''_{S_t S_t}(\diamond) = \partial_{S_t S_t}^2 \tilde{\Sigma}(\star) - \frac{2}{S_t} \partial_{S_t} \tilde{\Sigma}'_y(\star) + \frac{1}{S_t^2} \left[\tilde{\Sigma}'_{yy}(\star) + \tilde{\Sigma}''_{yy}(\star) \right]. \end{aligned} \quad (2.82)$$

We can now conclude our answer to the second question with the following remark:

Remark 2.11 Let us denote by (t, T) the *time* pair and (S_t, K) the *space* pair. Given the three functionals $\tilde{\Sigma}'_{\theta}(\star)$, $\tilde{\Sigma}'_y(\star)$ and $\tilde{\Sigma}''_{yy}(\star)$ then we have any cross-differential of the absolute (traded) price, provided one of the following *exclusive* conditions is verified:

- The cumulative differential orders of the time *and* space pairs are both 0.
- The cumulative differential order is 1 for the time and 0 for the space pair.
- The cumulative differential order is 1 or 2 for the space and 0 for the time pair.

These Greeks include the Theta, Delta, Gamma and by consequence the Vanna cross-differential (since the volatility differentials are available).

Note that the time vs space discrepancy is naturally linked to Itô's Lemma, and is manifest at all higher differential orders with the *ladder effect* (see Fig. 3.1, p. 122). Finally, recall that should we wish to go beyond the differential definition of the Greeks and assess the hedge by transferring the dynamics themselves, we can use (2.16) [p. 42] and (2.22) [p. 44].

2.5.2 Illustration: Qualitative Analysis of a Classical SV Model Class

Later on, we will dedicate a large part of Chap. 4 to practical applications of the direct results to current real-life models, such as SABR, thus showing how Theorem 2.2, along with the higher-order extensions found in Sect. 3.2, can be used to produce model-specific IATM differentials and whole-smile approximations.

In this section, our ambitions are more modest, and they also take a different angle. Indeed, we show how the first layer (Theorem 2.2) alone can be invoked, not to approach a whole marginal distribution, but to predict the rough *qualitative* features of a smile, when it is generated by a given stochastic instantaneous volatility model. In other words, we emphasise the potential of the asymptotic results as a rapid *analysis* toolbox for such SInSV models.

To that intent, we select a rich and very popular SV model class, the Extended Skew Market Model, which incorporates a local volatility component. We then examine the basic statics and dynamics of its implied volatility surface, focusing on the close relationship between the skew and the local volatility function. We will see that we can extend the more academic results of pure local volatility models, with little additional work.

We then specialise these asymptotic results, and show that they allow fast and easy comparison of two very popular local volatility instances: the Lognormal Displaced Diffusion (LDD) and the Constant Elasticity of Variance (CEV). We show that our asymptotic results can provide fast answers to qualitative questions regarding the respective generated smiles, in particular the IATM level, skew as well as their dynamics. Overall, we show also that the computations involved are simple and repetitive, which suggests a possible automation.

2.5.2.1 Properties of Local Volatility “Skew” Functions

So far we have mainly been considering generic formulae, with only a few errands in the local volatility framework. We have seen, for instance, in Remark 2.10 [p. 76] that, when designing or specifying the parameters of a stochastic *instantaneous* volatility model, the triplet $[\sigma_t, a_{2,t}, a_{22,t}]$ would be sufficient to simultaneously control the IATM level, the skew and the curvature.

Although it has proven efficient to build some useful intuition, that local volatility class is nevertheless limited as a real-life hedging model. Indeed, in practice such pure local volatility models allow us to match any smooth and valid smile, but exhibit poor, if not downright dangerous, dynamic properties (see [11] for instance). Essentially for that reason, both academics and practitioners have engineered a class extending the local volatility models, and which currently dominates the modelling approach. These modern market models, such as SABR or FL-SV, combine both *local* and *stochastic* volatility by using a multiplicative perturbation process: we will call them “Extended Skew Market Models” for reasons that will be made clear shortly. But first of all, let us formalise the generic model class:

Definition 2.2 (*Extended Skew Market Model*)

$$\frac{dS_t}{S_t} = \alpha_t f(t, S_t) dW_t \quad \text{or} \quad dS_t = \alpha_t \varphi(t, S_t) dW_t, \quad (2.83)$$

$$d\alpha_t = h(t, \alpha_t) dt + \varepsilon g(t, \alpha_t) dB_t \quad \text{with} \quad \rho dt = \langle dW_t, dB_t \rangle. \quad (2.84)$$

This class clearly represents a subset of the generic *parametric diffusion* model (2.79)–(2.80). We chose not to treat the latter here simply because of its (potentially) higher Markovian dimension and its more implicit correlation structure. Those generate formulae which are quite complex and therefore tend to be counter-productive for our modest purpose of illustration.

The ESMM class includes, among others, Heston [2], Lewis [1], SABR [11], FL-SV [50] and FL-TSS [51]: they all share its principle of a local volatility model which is then multiplicatively perturbed by the α_t process. Note, however, that in the literature the dynamics specified *in lieu* of (2.84) tend to be those of the “variance” α_t^2 rather than α_t proper.

As to why we chose to present both Normal and Lognormal conventions, it is simply because both are equally often found in the literature: for instance Lognormal is used with Heston, whereas Normal is used with SABR and FL-SV. The “skew model” terminology can now be justified: it comes from the fact that some of these models only allow zero correlation between the underlying and its perturbation process: this is the case for FL-SV, by opposition to SABR. When thus restricted to independent perturbations, they need rely purely on their local volatility function to generate some IATM skew, hence the name.

Conversely, note that this Extended Skew Model naturally incorporates the case of a deterministic perturbation $\alpha(t)$, hence falling back onto the pure local volatility models seen previously.

The desire to present results simultaneously for the Normal and Lognormal conventions of the ESMM leads us to first establish a very low-level transition formula. This is also justified by the fact that, throughout this study, we will frequently be juggling between Lognormal and Normal conventions, wherever dynamics are concerned. For instance, our results so far have been expressed using the Black implied volatility (hence the Lognormal “baseline”) while many models (including the ESMM class above) use Normal conventions instead. In Chap. 4, for example, high-order IATM differentials for SABR will require many such conversions. Another example of conversion between the Normal and Lognormal models can be found in Chap. 3, where baseline transfers are examined. For these reasons, establishing the following Lemma is clearly economical at the scale of the study, if a bit of an overkill for the current section.

Lemma 2.3 (Infinite-order differentiation of the Lognormal LV function) *Assume two $C^n, \mathbb{R} \rightarrow \mathbb{R}$ functions f and φ with $f(x) = x^{-1}\varphi(x)$. Then for $n \geq 0$,*

$$f^{(n)}(x) = \sum_{k=0}^n (-1)^k \frac{n!}{(n-k)!} x^{-(k+1)} \varphi^{(n-k)}(x).$$

Proof The result is clearly verified at the initial index $n = 0$.

Then by induction: we assume the property verified at index n , so that

$$\begin{aligned} f^{(n+1)} &= \sum_{k=0}^n \frac{(-1)^k n!}{(n-k)!} \left[(-1)(k+1)x^{-(k+2)} \varphi^{(n-k)}(x) + x^{-(k+1)} \varphi^{(n-k+1)}(x) \right] \\ &= \sum_{k=1}^{n+1} \frac{(-1)^k n!k}{(n-k+1)!} x^{-(k+1)} \varphi^{(n-k+1)}(x) + \sum_{k=0}^n \frac{(-1)^k n!}{(n-k)!} x^{-(k+1)} \varphi^{(n+1-k)}(x), \end{aligned}$$

hence

$$\begin{aligned} f^{(n+1)} &= x^{-1} \varphi^{(n+1)}(x) + (-1)^{n+1} (n+1)! x^{-(n+2)} \varphi(x) \\ &\quad + \sum_{k=1}^n (-1)^k \left[\frac{n!k(n-k)! + n!(n+1-k)!}{(n+1-k)!(n-k)!} \right] x^{-(k+1)} \varphi^{(n+1-k)}(x). \end{aligned}$$

The bracket on the right-hand side can then be rewritten as

$$\left[\frac{n!k(n-k)! + n!(n+1-k)(n-k)!}{(n+1-k)!(n-k)!} \right] = \left[\frac{(n+1)!}{(n+1-k)!} \right]$$

and we obtain the desired expression at index $n + 1$. □

In order to apply Theorem 2.2 the first step is generally to compute the native expressions for the coefficients of the SInsV model, as formatted in our framework by (2.8)–(2.10). Recall that those coefficients invoked in the first layer are only σ_t , $a_{1,t}$, $a_{2,t}$, $a_{3,t}$ and $a_{22,t}$.

Lemma 2.4 (Instantaneous coefficients of the ESMM) *Let us place ourselves within the Extended Skew Market Model defined by (2.83)–(2.84). In the following formulas, we assume the functions to be taken at the immediate point, i.e. at our current set of Markov variables (t, S_t, α_t) . Accordingly, we use the following notations:*

$$f \triangleq f(t, S_t) \quad \varphi \triangleq \varphi(t, S_t) \quad g \triangleq g(t, \alpha_t) \quad h \triangleq h(t, \alpha_t) \quad \dot{\rho} \triangleq \sqrt{1 - \rho^2}. \quad (2.85)$$

The instantaneous coefficients σ_t , a_2 , a_3 and a_{22} come as:

- Using the Lognormal convention for the ESMM:

$$\sigma_t = \alpha_t f \quad a_{2,t} = \sigma_t^2 S_t \frac{f'_2}{f} + \rho \varepsilon \sigma_t \frac{g}{\alpha_t} \quad a_{3,t} = \dot{\rho} \varepsilon \sigma_t \frac{g}{\alpha_t} \quad (2.86)$$

$$a_{1,t} = \sigma_t \frac{f'_1}{f} + \frac{1}{2} \sigma_t^3 S_t^2 \frac{f''_{22}}{f} + \sigma_t \frac{h}{\alpha_t} + \rho \varepsilon \sigma_t g S_t f'_2 \quad (2.87)$$

$$\begin{aligned} a_{22,t} = & \sigma_t^3 S_t \left[\frac{f'_2}{f} + S_t \left[\frac{f'_2}{f} \right]^2 + S_t \frac{f''_{22}}{f} \right] \\ & + 3 \rho \varepsilon \sigma_t g S_t f'_2 + \rho^2 \varepsilon^2 \sigma_t \frac{g}{\alpha_t} g'_2. \end{aligned} \quad (2.88)$$

- Using the Normal convention for the ESMM:

$$\sigma_t = \alpha_t S_t^{-1} \varphi \quad a_{2,t} = \sigma_t^2 \left[S_t \frac{\varphi'_2}{\varphi} - 1 \right] + \rho \varepsilon \sigma_t \frac{g}{\alpha_t} \quad a_{3,t} = \dot{\rho} \varepsilon \sigma_t \frac{g}{\alpha_t} \quad (2.89)$$

$$a_{1,t} = \sigma_t \frac{\varphi'_1}{\varphi} + \sigma_t^3 \left[1 - S_t \frac{\varphi'_2}{\varphi} + \frac{1}{2} S_t^2 \frac{\varphi''_{22}}{\varphi} \right] + \sigma_t \frac{h}{\alpha_t} + \rho \varepsilon \sigma_t g \left[\varphi'_2 - S_t^{-1} \varphi \right]$$

$$\begin{aligned} a_{22,t} = & \sigma_t^3 \left[2 - 3 S_t \frac{\varphi'_2}{\varphi} + S_t^2 \left[\frac{\varphi'_2}{\varphi} \right]^2 + S_t^2 \frac{\varphi''_{22}}{\varphi} \right] \\ & + 3 \rho \varepsilon g \sigma_t \left[\varphi'_2 - S_t^{-1} \varphi \right] + \rho^2 \varepsilon^2 \frac{g}{\alpha_t} g'_2 \sigma_t. \end{aligned} \quad (2.90)$$

Given the obvious filiation of the ESMM, it is natural, easy and reassuring to verify that these results match those obtained for pure LV models: see Lemma 2.2 [p. 68].

Proof (Lemma 2.4)

Using the Lognormal convention, the cast provides the instantaneous volatility:

$$\sigma_t = \alpha_t f(t, S_t). \quad (2.91)$$

Then, simply using Itô gives us its dynamics as

$$\begin{aligned} d\sigma_t = & \left[\alpha_t f'_1 + \frac{1}{2} \alpha_t f''_{22} \alpha_t^2 S_t^2 f^2 + \rho \varepsilon \alpha_t g S_t f f'_2 \right] dt \\ & + \alpha_t f'_2 \alpha_t S_t f dW_t + f h dt + f \varepsilon g \left[\rho dW_t + \sqrt{1 - \rho^2} dZ_t \right]. \end{aligned}$$

Furthermore, applying Lemma 2.3 for $n = 1$ and $n = 2$, we easily get the conversions

$$f'_2 = -S_t^{-2} \varphi + S_t^{-1} \varphi'_2 \quad \text{and} \quad f''_{22} = S_t^{-1} \varphi''_{22} - 2S_t^{-2} \varphi'_2 + 2S_t^{-3} \varphi. \quad (2.92)$$

Therefore, the first three instantaneous coefficients come respectively as

$$\begin{aligned} a_{1,t} &= \alpha_t f'_1 + \frac{1}{2} \alpha_t^3 S_t^2 f^2 f''_{22} + h f + \rho \varepsilon \alpha_t g S_t f f'_2 \\ &= \alpha_t S_t^{-1} \varphi'_1 + \frac{1}{2} \alpha_t^3 \varphi^2 \left[S_t^{-1} \varphi''_{22} - 2S_t^{-2} \varphi'_2 + 2S_t^{-3} \varphi \right] \\ &\quad + h S_t^{-1} \varphi + \rho \varepsilon \alpha_t g \varphi \left[-S_t^{-2} \varphi + S_t^{-1} \varphi'_2 \right] \end{aligned}$$

and

$$\begin{aligned} a_{2,t} &= \alpha_t^2 S_t f(t, S_t) f'_2(t, S_t) + \rho \varepsilon g(t, \alpha_t) f(t, S_t) \\ &= \alpha_t^2 \varphi(t, S_t) \left[-S_t^{-2} \varphi(t, S_t) + S_t^{-1} \varphi'_2(t, S_t) \right] + \rho \varepsilon g(t, \alpha_t) S_t^{-1} \varphi(t, S_t) \end{aligned}$$

and

$$a_{3,t} = \varepsilon \dot{\rho} g f = \varepsilon \dot{\rho} g S_t^{-1} \varphi.$$

We can then switch α_t and f with σ_t using (2.91), which provides the desired expressions. Note that we chose to conserve the arguments in the expression of the endogenous coefficient $a_{2,t}$, as we now need its dynamics:

$$\begin{aligned} da_{2,t} &= \left[\alpha_t^2 \left[f f'_2 + S_t f_2'^2 + S_t f f''_{22} \right] + \rho \varepsilon g f'_2 \right] \alpha_t S_t f dW_t \\ &\quad + \left[2S_t f f'_2 \alpha_t + \rho \varepsilon g'_2 f \right] \rho \varepsilon g dW_t + [\dots] dt + [\dots] dZ_t, \end{aligned}$$

where we again isolate the endogenous coefficient

$$a_{22,t} = \alpha_t^3 S_t f \left[f f_2' + S_t f_2'^2 + S_t f f_{22}'' \right] + 3\rho\varepsilon\alpha_t g S_t f f_2' + \rho^2 \varepsilon^2 g g_2' f.$$

Again switching σ_t with α_t and f in this expression gives the sought formula. We can also use the conversion formulas to obtain, under the Normal convention:

$$\begin{aligned} a_{22,t} = & 3\rho\varepsilon\alpha_t g \varphi \left[S_t^{-1} \varphi_2' - S_t^{-2} \varphi \right] + \rho^2 \varepsilon^2 g g_2' S_t^{-1} \varphi \\ & + \alpha_t^3 \varphi \left[S_t^{-1} \varphi \left[\frac{\varphi_2'}{S_t} - \frac{\varphi}{S_t^2} \right] + S_t \left[\frac{\varphi_2'}{S_t} - \frac{\varphi}{S_t^2} \right]^2 + \varphi \left[\frac{\varphi_{22}''}{S_t} - 2 \frac{\varphi_2'}{S_t^2} + 2 \frac{\varphi}{S_t^3} \right] \right], \end{aligned}$$

which after simplification and switch gives (2.90). $\square$

When performing this type of calculation, we recommend identifying as many recurring, symmetric or meaningful quantities as possible, as early as possible. Similarly, it is also a good idea to allocate *ad hoc* symbols to these quantities. It might not seem useful at the current level of the first layer, but when computing deeper coefficients (as needed by higher-order IATM differentials, see Sect. 3.2) they will come in handy. For examples of this “technique”, see, for instance, the treatment of SABR in Chap. 4.

We can now move on to the IATM differentials, both static and dynamic.

Proposition 2.3 (IATM differentials of the ESMM) *Within the ESMM framework, defining $(\star) \triangleq (t, 0, 0)$*

- *The IATM skew is expressed as*

$$\tilde{\Sigma}'_y(\star) = \tilde{\Sigma}'_{y,f}(\star) + \tilde{\Sigma}'_{y,\rho}(\star), \quad (2.93)$$

where

$$\tilde{\Sigma}'_{y,f}(\star) \triangleq \frac{1}{2} \sigma_t S_t \frac{f'}{f} = \frac{1}{2} \sigma_t \left[S_t \frac{\varphi_2'}{\varphi} - 1 \right]$$

is the “local volatility” term, while the “correlation” term comes as

$$\tilde{\Sigma}'_{y,\rho}(\star) \triangleq \frac{1}{2} \rho \varepsilon \frac{g}{\alpha_t}.$$

- *The IATM curvature is expressed as*

$$\tilde{\Sigma}''_{yy}(\star) = \tilde{\Sigma}''_{yy,f}(\star) + \tilde{\Sigma}''_{yy,\rho}(\star),$$

with

$$\tilde{\Sigma}_{yy,f}''(\star) \triangleq \frac{1}{3}\sigma_t \left[S_t \frac{f_2'}{f} - \frac{1}{2}S_t^2 \left[\frac{f_2'}{f} \right]^2 + S_t^2 \frac{f_{22}''}{f} \right]$$

$$\tilde{\Sigma}_{yy,\rho}''(\star) \triangleq \frac{\varepsilon^2}{3} \frac{1}{\sigma_t} \left[\rho^2 \frac{g}{\alpha_t} g' + \left(1 - \frac{5}{2}\rho^2 \right) \frac{g^2}{\alpha_t^2} \right].$$

In Normal representation, the “local volatility” term becomes

$$\tilde{\Sigma}_{yy,f}''(\star) = \frac{1}{6}\sigma_t \left[2S_t^2 \frac{\varphi_{22}''}{\varphi} - S_t^2 \left[\frac{\varphi_2'}{\varphi} \right]^2 + 1 \right].$$

- The IATM slope is expressed as

$$\tilde{\Sigma}'_{\theta}(\star) = \tilde{\Sigma}'_{\theta,f}(t, 0, 0) + \tilde{\Sigma}'_{\theta,\rho}(t, 0, 0) + \frac{\sigma_t}{4}\rho\varepsilon g S_t f'$$

with

$$\tilde{\Sigma}'_{\theta,f}(\star) \triangleq \frac{\sigma_t}{2} \frac{f_1'}{f} + \frac{\sigma_t^3}{12} \left[S_t \frac{f_2'}{f} - \frac{1}{2}S_t^2 \left[\frac{f_2'}{f} \right]^2 + S_t^2 \frac{f_{22}''}{f} \right]$$

$$\tilde{\Sigma}'_{\theta,\rho}(\star) \triangleq \frac{\sigma_t^2}{4}\rho\varepsilon \frac{g}{\alpha_t} + \frac{\sigma_t}{12}\varepsilon^2 \left[1 + \frac{1}{2}\rho^2 \right] \left[\frac{g}{\alpha_t} \right]^2 - \frac{\sigma_t}{6}\rho^2\varepsilon^2 \frac{g}{\alpha_t} g' + \frac{\sigma_t}{2} \frac{h}{\alpha_t}.$$

The local volatility term rewrites, in Normal convention, as:

$$\tilde{\Sigma}'_{\theta,f}(\star) = \frac{\sigma_t}{2} \frac{\varphi_1'}{\varphi} + \frac{\sigma_t^3}{24} \left[1 - S_t^2 \left[\frac{\varphi_2'}{\varphi} \right]^2 + 2S_t^2 \frac{\varphi_{22}''}{\varphi} \right].$$

- As for the dynamic coefficients, the IATM volatility of volatility is

$$\begin{cases} \tilde{v}(\star) = \sigma_t^2 S_t \frac{f_2'}{f} + \rho\varepsilon\sigma_t \frac{g}{\alpha_t} = \underbrace{\sigma_t^2 \left[S_t \frac{\varphi_2'}{\varphi} - 1 \right]}_{\tilde{v}_f(\star)} + \underbrace{\rho\varepsilon\sigma_t \frac{g}{\alpha_t}}_{\tilde{v}_\rho(\star)} \\ \tilde{n}(\star) = \dot{\rho}\varepsilon\sigma_t \frac{g}{\alpha_t} \end{cases}$$

while the endogenous volatility of the skew comes as

$$\tilde{v}'_y(\star) = \tilde{v}'_{y,f}(\star) + \tilde{v}'_{y,\rho}(\star) + \frac{1}{2}\rho\epsilon g S_t f'_2,$$

with

$$\tilde{v}'_{y,f}(\star) \triangleq \frac{1}{2}\sigma_t^2 \left[S_t \frac{f'_2}{f} + S_t^2 \frac{f''_{22}}{f} \right] = \frac{1}{2}\sigma_t^2 \left[1 - S_t \frac{\varphi'_2}{\varphi} + S_t^2 \frac{\varphi''_{22}}{\varphi} \right]$$

$$\tilde{v}'_{y,\rho}(\star) \triangleq \frac{1}{2}\rho^2\epsilon^2 \frac{g}{\alpha_t} \left[g'_2 - \frac{g}{\alpha_t} \right].$$

Instead of heading thereupon into the proof (which is very straightforward) let us first comment on and interpret these results; we start with general considerations, before reviewing some of the IATM differentials on an individual basis.

First of all, it seems rather intuitive that the formulae describing the ESMM shall prove more complex when using the Normal convention. Indeed, recall that the generic dynamics for S_t , which define σ_t and therefore a_1 , a_2 , a_3 and a_{22} , are Lognormal; likewise, the implied volatility $\tilde{\Sigma}$ that we consider is also Lognormal. In other words, we have on one hand a “simple” (Lognormal) model used to define the implied volatility—the *baseline*—and on the other hand, a “complex” target model written along Normal lines: this dissimilarity is what complexifies the results. In Chap. 3, we will see that this also impacts the precision and usefulness of the asymptotic method. Nevertheless, we will also show that adapting the baseline to the target model is often possible. In particular the Normal baseline is frequently a natural option, and this would certainly be the case here.

Still on the presentation of those results, the main reason why we choose to position the differentials of f as numerators is to underline the inherent *scaling* of the problem. In other words, it stresses the fact that the IATM differentials correspond to *shape* factors of the local volatility function, while only the level depends on its scale.

The relative complexity of the formulas, in conjunction with the scaling aspect, point naturally to a change of variables. This is indeed no surprise, since the left-hand-sides of these equations represent (differentials of) functions of log-moneyness $\ln(K/S_t)$, while the right-hand sides are functions of S_t . We will use this feature in the sequel to rewrite some of the quantities in a more intuitive manner.

Evidently, it is easy to see that by neutralising the perturbation we fall back onto the expressions for pure local volatility models: refer to (2.67) for the level, (2.69) for the skew, (2.72) for the curvature, and (2.74) for the slope. Accordingly we find that, as was the case with pure local volatility models, once the perturbation process α_t is set then the differentials f , f'_2 , f''_{22} and f'_1 taken in (t, S_t) respectively control the IATM level, skew, curvature and slope.

But the fundamental feature to note, in all the IATM differentials expressions for the ESMM, is their clean decomposition between local volatility and correla-

tion terms. Indeed, cross-terms only appear in the expressions for $\tilde{\Sigma}'_{\theta}(\star)$ and $\tilde{V}'_y(\star)$. Moreover, these two terms are very similar and can easily be interpreted further, as will be discussed shortly. This decomposition implies that we can interpret many features of the model by referring to, respectively, pure local volatility models (Dupire) or pure correlation models (such as Heston). These two classes are extreme instances of the ESMM: the local volatility model is obtained with $\alpha_t \equiv 1$ or $\sigma_t \equiv f(t, S_t)$, while the pure correlation model corresponds to $f(t, S_t) \equiv 1$ or $\sigma_t \equiv \alpha_t$.

Let us now comment individually on the IATM differentials. First, with respect to the skew $\tilde{\Sigma}'_y(\star)$, we remark that the correlation term $\tilde{\Sigma}'_{y,\rho}$ is by definition unaffected by the presence, and therefore choice, of the local volatility function $f(\cdot)$. Consequently, when considering *skew vs local volatility* issues in this model framework, we will focus purely on $\tilde{\Sigma}'_{y,f}$, which is equivalent to considering pure local volatility models.

That said, the correlation parameter does provide an additional degree of freedom which helps control the IATM skew. The Heston model for instance, relies solely on this correlation to create the descriptor, but can nevertheless be calibrated to fairly extreme market skews.

In fact, we observe it is actually the term $\rho \varepsilon g$, i.e. *the product of correlation with the Lognormal volatility of the perturbation*²⁸ which will provide that additional skew. In other words, the vol of vol “activates” and “compounds” the correlation, which is not to say that they cannot be dissociated, as proven by the terms $\tilde{\Sigma}''_{yy,\rho}$ and $\tilde{\Sigma}'_{\theta,\rho}(\star)$. Overall, this behaviour should come as no surprise to any modeller or practitioner with experience of Heston or SABR, for instance.

Conversely, when the correlated perturbation α_t is not available (as per FL-SV [50] or FL-TSS [51], which both use an *independent* perturbation), the skew function $f(\cdot)$ might have to be fairly extreme to match the market skews. It is possible to use a very steep $f(\cdot)$ around a certain point S_t^* , but not across a whole range of underlying's values. This means that the *conditional smile* will lose its skew, whereas there is no such constraint with $\tilde{\Sigma}'_{y,\rho}(\star)$, which can consistently create skew for all values of S_t . Not surprisingly, there is a strong analogy with the smile curvature issue (local vs stochastic volatility) as discussed in [11].

In terms of calibration capacity, this is a main advantage of the SABR class, for instance. Indeed, this model can use correlation to control its skew, allowing us to specify the LV function to achieve other aims, such as controlling the backbone.

Turning now to the curvature $\tilde{\Sigma}''_{yy}(\star)$, we observe the same decomposition between the effects of local volatility and correlation. This property is very convenient, both for model design and for static calibration. The relative complexity of the $\tilde{\Sigma}''_{yy,f}(\star)$ term suggests aligning the variables on both sides: indeed, we have a differential w.r.t. log-moneyness y on the left-hand side, and w.r.t. S_t on the right-hand side. We elect to define a new local volatility function in sliding coordinates *via*

$$l(t, y) \triangleq f(t, S_t).$$

²⁸ $\varepsilon \alpha_t^{-1} g(\alpha_t)$, sometimes abusively called “vol of vol”.

Then the transition formulae of Appendix B allow us to rewrite

$$\begin{aligned}\tilde{\Sigma}_{yy,f}''(\star) &= \frac{1}{3}\sigma_t \left[S_t \frac{1}{l} \left[\frac{-1}{S_t} \right] l_2' - \frac{1}{2} S_t^2 \left[\frac{1}{l} \left[\frac{-1}{S_t} \right] l_2' \right]^2 + S_t^2 \frac{1}{l} \frac{1}{S_t^2} [l_2' + l_{22}''] \right] \\ &= \frac{1}{3}\sigma_t \left[\frac{l_{22}''}{l} - \frac{1}{2} \left[\frac{l_2'}{l} \right]^2 \right],\end{aligned}$$

which gains in coherence and compactness, but also lends itself to an easier interpretation. Note, in particular, that the second term corresponds to the squared skew, and also that the scaling property has been maintained. Indeed, the equation links the normalised curvature $\sigma_t^{-1} \tilde{\Sigma}_{yy}''(\star)$ to the *shape* of l , and not its scale.

Looking at the expression for $\tilde{V}_y'$, notwithstanding a single cross term, we observe a similar decomposition between local volatility and correlation terms. Focusing first on the local volatility term, it appears that it corresponds to the endogenous volatility of the LV term of the skew. In other words, the LV- ρ split extends consistently to the dynamics, just as if we were dealing with a pure LV model:

$$\begin{aligned}d\tilde{\Sigma}_y'(\star) &\stackrel{\text{LV}}{\equiv} d\tilde{\Sigma}_{y,f}'(\star) = [\cdot] dt + \frac{1}{2} [f_2' + S_t f_{22}''] dS_t = [\cdot] dt + \tilde{V}_{y,f}'(\star) dW_t \\ \text{with } \tilde{V}_{y,f}'(\star) &\triangleq \frac{1}{2}\sigma_t [S_t f_2' + S_t^2 f_{22}''].\end{aligned}$$

To reduce the complexity of the $\tilde{V}_{y,f}'(\star)$ term, we can proceed in the same manner as before by changing to sliding variables. The new expression reads as

$$\tilde{V}_{y,f}'(\star) = \frac{1}{2}\sigma_t^2 \left[S_t \left(-\frac{1}{S_t} \right) \frac{l_2'}{l} + S_t^2 \frac{1}{S_t^2} \left[\frac{l_2' + l_{22}''}{l} \right] \right] = \frac{1}{2}\sigma_t^2 \frac{l_{22}''}{l}.$$

In other words, the convexity of the local volatility function creates positively correlated dynamics for the skew (and conversely if concave).

As for the correlation term $\tilde{V}_{y,\rho}'(\star)$, we again observe that the split between LV and ρ is extended to the dynamics, as if we were dealing with a pure correlation model such as Heston:

$$\begin{aligned}d\tilde{\Sigma}_y'(\star) &\stackrel{\text{Pure } \rho}{\equiv} d\tilde{\Sigma}_{y,\rho}'(\star) = [\cdot] dt + \frac{1}{2}\rho\varepsilon \left[\frac{g_2'}{\alpha_t} - \frac{g}{\alpha_t^2} \right] d\alpha_t = [\cdot] dt + \tilde{V}_{y,\rho}'(\star) dW_t \\ \text{with } \tilde{V}_{y,\rho}'(\star) &\triangleq \frac{1}{2}\rho^2\varepsilon^2 \frac{g}{\alpha_t} \left[g_2' - \frac{g}{\alpha_t} \right].\end{aligned}$$

In the $\tilde{V}_{y,\rho}'(\star)$ expression itself, we can interpret the bracket as representing the local “over-linearity” of the vol of vol function g . In practice, the function g is chosen so that

$g(0) = 0$ in order to maintain the non-negativity of the multiplicative perturbation α_t . It is also often taken as a power: typically $g(x) = x^\beta$ with $\beta \leq 1$ to ensure sub-linear growth and therefore existence of the solution for the SDE (2.84) (as per Itô's conditions). This is in particular the case for Heston ($\beta = -1$), SABR ($\beta = 1$) and most FL-SV or FL-TSS implementations (if $\psi(V) = V^\gamma$ for the variance then $\beta = 2\gamma - 1$). In that case we have the correlation term

$$\tilde{v}'_{y,\rho}(\star) = \frac{1}{2}\rho^2\varepsilon^2\alpha_t^{\beta-1}[\beta - 1]\alpha_t^{\beta-1} \leq 0. \quad (2.94)$$

This formula tells us that whenever g is a sub-linear power, the contribution of the effective correlation (i.e. *the product* $\rho\varepsilon g$) on the skew's endogenous volatility will be *negative*. In other words, it will tend to make the skew and the asset negatively correlated.

It would certainly be tempting to assume that an upward movement of the underlying S_t will (on average) be associated to a downward move of the IATM skew $\tilde{\Sigma}'_y(\star)$. Unfortunately at this stage of the book we do not possess the required results to conclude so: in contrast to the underlying, the IATM skew *does* exhibit a drift coefficient $\tilde{b}'_y(\star)$, and this finite variation term has *a priori* no reason to be null.

It would be an easy task to derive the expression for $\tilde{b}'_y(\star)$ in the simple Heston framework for instance, but this falls outside the scope of this simple interpretative section. The generic formula, however, will be provided in Sect. 3.2: see (E.30) [p. 449].

Finally, let us link and interpret the two cross-terms appearing in the expressions of $\tilde{\Sigma}'_\theta(\star)$ and $\tilde{v}'_y(\star)$. They are indeed almost identical, and can usefully be linked to the local volatility and correlation terms of the IATM skew, by noticing that

$$\rho\varepsilon g(t, \alpha_t)S_t f'_2(t, S_t) = 4\tilde{\Sigma}'_{y,f}(\star)\tilde{\Sigma}'_{y,\rho}(\star),$$

which explains why we did not deem it useful to provide either cross-term with Normal conventions. Having completed this interpretative review, let us now move on to the proof.

Proof (Proposition 2.3) Using the generic result (2.51) and the expressions for σ_t and a_2 as per (2.86) the skew comes as

$$\tilde{\Sigma}'_y = \frac{1}{2}\sigma_t S_t \frac{f'_2}{f} + \frac{1}{2}\rho\varepsilon \frac{g}{\alpha_t} = \frac{1}{2}\sigma_t \left[S_t \frac{\varphi'_2}{\varphi} - 1 \right] + \frac{1}{2}\rho\varepsilon \frac{g}{\alpha_t}.$$

For future use, we pre-compute

$$a_{2,t}^2 = \left[\sigma_t^2 S_t \frac{f'_2}{f} + \rho\varepsilon \sigma_t \frac{g}{\alpha_t} \right]^2 = \sigma_t^4 S_t^2 \left[\frac{f'_2}{f} \right]^2 + \rho^2 \varepsilon^2 \sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 + 2\rho\varepsilon \sigma_t^2 g S_t f'_2.$$

Then using the generic curvature formula (2.52) along with the specific expressions for a_3 and a_{22} given respectively by (2.86) and (2.88), the ESMM curvature comes as

$$\begin{aligned}\tilde{\Sigma}_{yy}''(\star) = & \frac{1}{3\sigma_t^2} \left[\sigma_t^3 S_t \left[\frac{f_2'}{f} + S_t \left[\frac{f_2'}{f} \right]^2 + S_t \frac{f_{22}''}{f} \right] + 3\rho\varepsilon\sigma_t g S_t f_2' + \rho^2\varepsilon^2\sigma_t \frac{g}{\alpha_t} g_2' \right] \\ & + \frac{1-\rho^2}{3\sigma_t^3} \varepsilon^2 \sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 - \frac{1}{2\sigma_t^3} \left[\sigma_t^4 S_t^2 \left[\frac{f_2'}{f} \right]^2 + \rho^2\varepsilon^2\sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 + 2\rho\varepsilon\sigma_t^2 g S_t f_2' \right],\end{aligned}$$

which, after simplifying and grouping the terms, yields

$$\tilde{\Sigma}_{yy}''(\star) = \frac{\sigma_t}{3} \left[S_t \frac{f_2'}{f} - \frac{1}{2} S_t^2 \left[\frac{f_2'}{f} \right]^2 + S_t^2 \frac{f_{22}''}{f} \right] + \frac{\varepsilon^2}{3\sigma_t} \left[\rho^2 \frac{g}{\alpha_t} g_2' + \left[1 - \frac{5}{2}\rho^2 \right] \frac{g^2}{\alpha_t^2} \right],$$

where only the first term depends on the local volatility function. With the Normal convention, this term becomes

$$\begin{aligned}\tilde{\Sigma}_{yy,f}''(\star) = & \frac{1}{3}\sigma_t \left[S_t \left[\frac{\varphi_2'}{\varphi} - \frac{1}{S_t} \right] - \frac{1}{2} S_t^2 \left[\frac{\varphi_2'}{\varphi} - \frac{1}{S_t} \right]^2 + S_t^2 \left[\frac{\varphi_{22}''}{\varphi} - \frac{2}{S_t} \frac{\varphi_2'}{\varphi} + \frac{2}{S_t^2} \right] \right] \\ = & \frac{1}{6}\sigma_t \left[2S_t^2 \frac{\varphi_{22}''}{\varphi} - S_t^2 \left[\frac{\varphi_2'}{\varphi} \right]^2 + 1 \right].\end{aligned}$$

Let us now tackle the slope: substituting (2.86), (2.87) and (2.88) into (2.53), we get

$$\begin{aligned}\tilde{\Sigma}_{\theta}'(\star) = & \frac{\sigma_t}{4} \left[\sigma_t^2 S_t \frac{f_2'}{f} + \rho\varepsilon\sigma_t \frac{g}{\alpha_t} \right] + \frac{1}{12\sigma_t} (1-\rho^2)\varepsilon^2 \sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 \\ & + \frac{1}{2} \left[\sigma_t \frac{f_1'}{f} + \frac{1}{2} \sigma_t^3 S_t^2 \frac{f_{22}''}{f} + \sigma_t \frac{h}{\alpha_t} + \rho\varepsilon\sigma_t g S_t f_2' \right] \\ & - \frac{1}{6} \left[\sigma_t^3 S_t \left[\frac{f_2'}{f} + S_t \left[\frac{f_2'}{f} \right]^2 + S_t \frac{f_{22}''}{f} \right] + 3\rho\varepsilon\sigma_t g S_t f_2' + \rho^2\varepsilon^2\sigma_t \frac{g}{\alpha_t} g_2' \right] \\ & + \frac{1}{8\sigma_t} \left[\sigma_t^4 S_t^2 \left[\frac{f_2'}{f} \right]^2 + \rho^2\varepsilon^2\sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 + 2\rho\varepsilon\sigma_t^2 g S_t f_2' \right].\end{aligned}$$

After simplification and grouping the terms, we obtain

$$\begin{aligned}\widetilde{\Sigma}'_{\theta}(\star) &= \underbrace{\frac{\sigma_t}{2} \frac{f'_1}{f} + \frac{\sigma_t^3}{12} \left[S_t \frac{f'_2}{f} - \frac{1}{2} S_t^2 \left[\frac{f'_2}{f} \right]^2 + S_t^2 \frac{f''_{22}}{f} \right]}_{\widetilde{\Sigma}'_{\theta,f}(\star)} + \frac{\sigma_t}{4} \rho \varepsilon g S_t f' \\ &\quad + \frac{\sigma_t^2}{4} \rho \varepsilon \frac{g}{\alpha_t} + \frac{\sigma_t}{12} \varepsilon^2 \left[1 + \frac{1}{2} \rho^2 \right] \left[\frac{g}{\alpha_t} \right]^2 - \frac{\sigma_t}{6} \rho^2 \varepsilon^2 \frac{g}{\alpha_t} g' + \frac{\sigma_t}{2} \frac{h}{\alpha_t}.\end{aligned}$$

Let us convert the first term using the Normal convention:

$$\begin{aligned}\widetilde{\Sigma}'_{\theta,f}(\star) &= \frac{\sigma_t}{2} \frac{\varphi'_1}{\varphi} + \frac{\sigma_t^3}{12} \left[S_t \frac{\varphi'_2 - S_t^{-1} \varphi}{\varphi} - \frac{1}{2} \left[S_t \frac{\varphi'_2 - S_t^{-1} \varphi}{\varphi} \right]^2 \right. \\ &\quad \left. + S_t^2 \left[\frac{\varphi''_{22} - 2S_t^{-1} \varphi'_2 + 2S_t^{-2} \varphi}{\varphi} \right] \right],\end{aligned}$$

which after simplification and grouping provides the desired result. Turning at last to the dynamic coefficients, we substitute (2.86) and (2.88) into (2.56) to obtain

$$\begin{aligned}\widetilde{v}'_y(\star) &= -\frac{1}{2\sigma_t^2} \left[\sigma_t^4 S_t^2 \left[\frac{f'_2}{f} \right]^2 + \rho^2 \varepsilon^2 \sigma_t^2 \left[\frac{g}{\alpha_t} \right]^2 + 2\rho \varepsilon \sigma_t^2 g S_t f'_2 \right] \\ &\quad + \frac{1}{2\sigma_t} \left[\sigma_t^3 S_t \left[\frac{f'_2}{f} + S_t \left[\frac{f'_2}{f} \right]^2 + S_t \frac{f''_{22}}{f} \right] + 3\rho \varepsilon \sigma_t g S_t f'_2 + \rho^2 \varepsilon^2 \sigma_t \frac{g}{\alpha_t} g'_2 \right],\end{aligned}$$

which, after simplification, yields

$$\widetilde{v}'_y(\star) = \underbrace{\frac{1}{2} \sigma_t^2 \left[S_t \frac{f'_2}{f} + S_t^2 \frac{f''_{22}}{f} \right]}_{\widetilde{v}'_{y,f}(\star)} + \frac{1}{2} \rho^2 \varepsilon^2 \frac{g}{\alpha_t} \left[g'_2 - \frac{g}{\alpha_t} \right] + \frac{1}{2} \rho \varepsilon g S_t f'_2$$

with

$$\begin{aligned}\widetilde{v}'_{y,f}(\star) &= \frac{1}{2} \sigma_t^2 \left[-S_t \frac{\varphi'_2 - S_t^{-1} \varphi}{\varphi} + S_t^2 \frac{\varphi''_{22} - 2S_t^{-1} \varphi'_2 + 2S_t^{-2} \varphi}{\varphi} \right] \\ &= \frac{1}{2} \sigma_t^2 \left[1 - S_t \frac{\varphi'_2}{\varphi} + S_t^2 \frac{\varphi''_{22}}{\varphi} \right],\end{aligned}$$

which concludes the proof. $\square$

2.5.2.2 Application: Comparison of the LDD and CEV Skew Functions

We have established with Proposition 2.3 the Layer-1 results for the generic “Extended Skew” class: we apply them now to the two most popular forms of local volatility functions, the Lognormal Displaced Diffusion (LDD) and the Constant Elasticity of Variance (CEV). These two models are commonly and concurrently used in their pure local volatility version (see [52] for a comparison), i.e. with α_t as a deterministic function of time. We can easily extend them by adding a multiplicative perturbation as per (2.83)–(2.84) so that they fit into the ESMM framework. Following the Normal convention, the models are then specified with:

$$\text{LDD : } \varphi_d(s, x) = \lambda_d(s) (x + d(s))$$

$$\text{CEV : } \varphi_\beta(s, x) = \lambda_\beta(s) x^{\beta(s)}.$$

Both forms of local volatilities are commonly used, either on a standalone basis or within two-dimensional local/stochastic volatility models (such as SABR). Their role is essentially to simultaneously control the static IATM level and *skew*, exploiting the fact that each is specified by a pair of time functions: $\lambda_d(\cdot)$ and $d(\cdot)$ vs $\lambda_\beta(\cdot)$ and $\beta(\cdot)$. However, either local volatility function inevitably generates some unwanted and rather distinct side-effects. So that in the literature, as well as among practitioners, the argument for and against these two respective forms of skew functions has been a recurring issue of the modelling process. Rather surprisingly, few papers are available that provide a clean answer to these questions. For a prompt review of CEV properties (which is more technical than the LDD) a good reference is [53], while a comparison of the two models (based on a small-time expansion) is provided by [52].

In this context, we will show that the asymptotic results of Theorem 2.2 are well suited to settle (almost) effortlessly a large part of that old argument. Our objective is to justify rigorously how to achieve the joint level and skew fit, and in particular how to match one model onto the other. Naturally this will expose the limits of such an approach for the smile, both in static and dynamic terms. In order to ensure comparability, we assume *the same multiplicative perturbation process* α_t .

We show that once the static level/skew is achieved, the sliding IATM option will have the same volatility (both endogenous and exogenous coordinates) in both cases, and therefore that the fixed-strike IATM option will have the same Delta. However, we will also see that other important quantities will necessarily be different, and how to approximate these discrepancies. Of particular interest for trading are the static smile IATM curvature, the volatility of the skew, the drift of the sliding IATM option, and the Gamma of the fixed-strike IATM option.

As we have seen with Proposition 2.3, with regard to IATM level and skew the effects of correlation (and thus of the perturbation α_t) are dissociated from those of the local volatility. Therefore it makes sense to recall the structural differences between the two models in their pure LV form. It is well-known that when d and β are either constant or time functions, both pure local volatility models provide closed-form solutions for Calls and Puts, which is a large part of their appeal. However

the solutions of their respective SDEs exhibit significantly distinct behaviours. Of particular concern are the existence and uniqueness of the CEV solution²⁹ and the support of the distribution for the LDD solution $([-d, +\infty[)$. Also, the (right) tail characteristics can be very different, affecting several quasi-vanilla products, such as CMS options.

The extension of the two models to a diffusive α_t inherits and enhances most of these differences. Furthermore, the ESMM decomposition between LV and perturbation effects ensures that few novel features are introduced. For instance, the backbone is controlled mainly by the LV function³⁰ and will therefore produce distinct *conditional* smiles, although the presence of stochastic volatility does ensure *some* Delta stickiness. Although the joint local/stochastic volatility context of the ESMM allows for further control of the smile properties, we will (artificially) consider that the primary role of $\varphi(\cdot, \cdot)$ is to control the IATM level and skew.

Our first move is obviously to unify the two local functions under the Extended Displaced CEV denomination, which is discussed for instance (in a time-homogeneous version) within [53].

Definition 2.3 (*The Extended Displaced CEV model*) Consider the EDCEV model to be specified as follows:

$$\begin{aligned} dS_t &= \alpha_t \varphi_{d,\beta}(t, S_t) dW_t & \text{with} & \quad \varphi_{d,\beta}(s, x) = \lambda(s)(x + d)^\beta \\ d\alpha_t &= h(t, \alpha_t)dt + \varepsilon g(t, \alpha_t)dB_t & \text{with} & \quad \langle dW_t, dB_t \rangle = \rho dt. \end{aligned} \quad (2.95)$$

We can then apply Proposition 2.3 directly in order to obtain the relevant IATM differentials. Again, due to the clear distinction between the correlation and local volatility effects, it is sufficient to express only the latter.

Corollary 2.12 (IATM level and skew of the EDCEV model) *Defining $(\star) \triangleq (t, 0, 0)$ the IATM level and skew are*

$$\tilde{\Sigma}(\star) = \alpha_t \lambda(t) S_t^{-1} (S_t + d)^\beta \quad \tilde{\Sigma}'_{y,f}(\star) = \frac{1}{2} \tilde{\Sigma}(\star) \left[\frac{\beta(t) S_t}{S_t + d(t)} - 1 \right], \quad (2.96)$$

which we can also express in strike units, using Corollary 2.5:

$$\Sigma'_{K,f}(LDD) = -\frac{1}{2} \frac{\sigma_t}{S_t} \frac{d}{S_t + d} \quad \Sigma'_{K,f}(CEV) = -\frac{1}{2} \frac{\sigma_t}{S_t} (1 - \beta).$$

Consider CEV and LDD models with matching IATM levels and skew.

²⁹ For positive β : if $\beta < 1$ zero is an attainable boundary, for $\frac{1}{2} \leq \beta \leq 1$ it is an absorbing boundary, and for $0 < \beta < \frac{1}{2}$ there is no uniqueness of the solution unless a boundary condition is specified (absorption or reflection). Obviously for $\beta > 1$ the sub-linear growth Itô condition ensuring existence is not satisfied.

³⁰ $\tilde{\Sigma}(\star) = \alpha_t S_t^{-1} \varphi(S_t)$.

Ignoring the IATM argument for $\tilde{\Sigma}$ and $\tilde{\Sigma}'_{y,f}$, we provide the transfer formulas:

$$\lambda_d = \lambda_\beta \beta S_t^{\beta-1} = \frac{1}{\alpha_t} \left[2\tilde{\Sigma}'_{y,f} + \tilde{\Sigma} \right] \quad \text{and} \quad d = \frac{1-\beta}{\beta} = \frac{-2S_t \tilde{\Sigma}'_{y,f}}{2\tilde{\Sigma}'_{y,f} + \tilde{\Sigma}},$$

$$\lambda_\beta = \lambda_d (S_t + d) S_t^{-\frac{S_t}{S_t+d}} = \frac{1}{\alpha_t} \tilde{\Sigma} S_t^{2\tilde{\Sigma}^{-1} \tilde{\Sigma}'_{y,f}} \quad \text{and} \quad \beta = \frac{S_t}{S_t + d} = 1 + 2\tilde{\Sigma}^{-1} \tilde{\Sigma}'_{y,f}.$$

Note that by normalising the displacement with $d(t) = c(t)S_t$ the equivalence between the two skew parameters becomes even more obvious. Indeed, the local volatility-induced IATM skew is then

$$\tilde{\Sigma}'_{y,f}(t, 0, 0) = \frac{1}{2} \tilde{\Sigma}(\star) \left[\frac{\beta(t)}{1 + c(t)} - 1 \right]. \quad (2.97)$$

Formulae (2.96) and (2.97) show that the E-LDD and E-CEV models can *a priori* be matched simultaneously in level and in (negative) skew. The latter is clearly bounded, even more so in the pure LDD and CEV cases, where for positive $\beta(\cdot)$ and $c(\cdot)$ the value of $-1/2 \tilde{\Sigma}(\star)$ cannot be exceeded (see [54] (p. 5) for instance). However, each model achieves these targets with distinct consequences on the marginal distributions. For example, recall that a CEV with $\beta = 1/3$ demands specification of the boundary conditions, while the equivalent LDD requires a displacement of $d = 2S_t$ to achieve the same skew: for most underlyings this support is considered to be unrealistic. By comparison, the EDCEV model allows us to reach the same negative skews in a healthier way, for instance by taking $\beta = 1/2$ and then using the displacement to complete. Overall, these results expose the limitations of pure local volatility models vs their (correlated) stochastic volatility extensions. Let us now move on to the proof.

Proof In this proof we will omit the arguments of $\tilde{\Sigma}$ and $\tilde{\Sigma}'_{y,f}$, which will be taken at $(t, 0, 0)$. Within (2.96) the IATM level is given by the cast, while the IATM skew comes directly from (2.93). Let us first consider an Extended LDD model with a given perturbation process α_t , and assume that it is matched to a given IATM level and skew. Then the skew match gives us

$$\tilde{\Sigma}'_{y,f} = \frac{1}{2} \tilde{\Sigma} \left[\frac{S_t}{S_t + d} - 1 \right] \quad \text{hence} \quad \frac{S_t}{S_t + d} = \frac{2\tilde{\Sigma}'_{y,f} + \tilde{\Sigma}}{\tilde{\Sigma}} \quad \text{thus} \quad d = \frac{-2S_t \tilde{\Sigma}'_{y,f}}{2\tilde{\Sigma}'_{y,f} + \tilde{\Sigma}},$$

while the level match provides

$$\tilde{\Sigma} = \alpha_t \lambda_d S_t^{-1} (S_t + d) \quad \text{hence} \quad \lambda_d = \frac{1}{\alpha_t} \tilde{\Sigma} \frac{S_t}{S_t + d} = \frac{1}{\alpha_t} \left[2\tilde{\Sigma}'_{y,f} + \tilde{\Sigma} \right].$$

Let us now assume instead that this Extended LDD is matched to a given Extended CEV model. Then the skew match followed by the level match give us sequentially

$$\frac{1}{S_t + d} = \frac{\beta}{S} \Rightarrow d = \frac{1 - \beta}{\beta} \quad \text{then} \quad \lambda_d [S_t + d] = \lambda_\beta S^\beta \Rightarrow \lambda_d = \beta \lambda_\beta S_t^{\beta-1}.$$

Conversely, let us then assume a CEV model, and suppose that it is matched to a given IATM level and skew. Then matching the skew gives us

$$\tilde{\Sigma}'_{y,f} = \frac{1}{2} \tilde{\Sigma} [\beta - 1] \quad \text{hence} \quad \beta = 1 + 2 \tilde{\Sigma}^{-1} \tilde{\Sigma}'_{y,f},$$

while the level match gives

$$\tilde{\Sigma} = \alpha_t \lambda_\beta S_t^{\beta-1} \Rightarrow \lambda_\beta = \frac{1}{\alpha_t} \tilde{\Sigma} S_t^{2\tilde{\Sigma}^{-1} \tilde{\Sigma}'_{y,f}}. \quad (2.98)$$

Suppose now that this Extended CEV is matched to a given Extended LDD model. Inverting the skew result in (2.96) before matching the level provides sequentially

$$\beta = \frac{S_t}{S_t + d} \quad \text{then} \quad \lambda_\beta = \lambda_d (S_t + d) S_t^{-\beta} = \lambda_d (S_t + d) S_t^{-\frac{S_t}{S_t + d}},$$

which concludes the proof. $\square$

We have already noted in Sect. 2.3.2 that if two stochastic volatility models (whether SinsV or SImpV) provide matching IATM level and skew, then their endogenous volatilities a_2 (or $\tilde{v}(\star)$) are identical. In the general case, it is not true, however, that the total vol of vol $[a_2; a_3]$ or even its modulus $\sqrt{a_2^2 + a_3^2}$ will be the same. As shown by the expression for a_3 within (2.89) however, this is the case for the Extended Skew Market Model, and in particular for the Extended LDD and CEV. This is important in terms of trading because these dynamics are those of the sliding IATM implied volatility and will therefore impact the *Delta* of the *fixed-strike* ATM Call. Indeed, as discussed in Sect. 2.5.1.4 and according in particular to (2.82), that Delta will be identical if both models are *dynamically* matched in level and in skew.

However, the lack of additional control parameters means that the above quantities represent the extent of any rigorous, IATM match between the two models. Indeed, differences will appear as soon as φ''_{22} or f''_{22} is invoked. In terms of instantaneous coefficients, this concerns a_1 and a_{22} , while in terms of IATM differentials it regards:

- the IATM curvature $\tilde{\Sigma}''_{yy}(\star)$;
- the slope $\tilde{\Sigma}'_\theta(\star)$;
- the endogenous volatility of the skew $\tilde{v}'_y(\star)$;
- the Gamma of the fixed-strike ATM option.

Overall, these are IATM results and a natural question to ask is whether the match of the EDCEV model is also possible for $\theta > 0$. The answer is that in general we can match ATM level and skew at a given maturity, because we have two targets (level and skew) and also two control parameters at our disposal. However, mathematically we cannot guarantee that this system of equations admits a (unique) solution.

2.5.3 Second Illustration: Smile-Specification of SInsV Models

In this section, we take the reverse approach from the previous illustration: starting from a given smile, we show how to build a stochastic instantaneous volatility model that would (approximatively) generate that surface. Taking that approach further, we make the case for the re-parametrisation of such models into what we call their “intuitive” versions. The latter consist in replacing one, several or all instantaneous parameters by more “trading-friendly”, smile-related quantities. First we discuss some generalities relevant to that approach, before exposing several examples.

2.5.3.1 Intuitive Models: Principles

An intuitive modelling approach consists in an output-oriented re-parametrisation of an existing SinsV model. The idea is to expose that original model to the practitioner (typically a trader) by using parameters that relate to *observable* and *meaningful* quantities. The first qualifier limits us to liquid instruments, whose characteristics are either quoted (e.g. prices, rates, implied volatilities or correlations) or easily measurable (e.g. realised volatility, historical correlation). The second qualifier pertains to those quantities which have an actual influence on the agent’s portfolio: they impact the prices, the hedges and/or the risk. Therefore, in order to link to the usual pricing, hedging and risk engines, we still need to internally express the native parameters driving the original SDE system.

Simple illustrations of this approach can already be found in practice-oriented literature. A classical instance consists in re-expressing volatility mean-reversion, in Heston’s model for instance, as “frequencies” or time periods in the one-dimensional dynamics of the volatility. Indeed, it is difficult to gauge the impact of selecting, say, $\kappa = 0.583$, and especially so because of the interaction of that parameter with the vol of vol and the current level of volatility. It seems more practical however, and particularly for traders, to consider the associated (i.e. model-implied) half-life of an after-shock decay: because that quantity can readily be compared to historical data, simply by looking at a realised chart of ATM implied volatility.

Going back to the general approach, what our asymptotic results allow us is to extend and generalise this approach to a much larger scale, to very complex models, while keeping some useful flexibility. The market quantity that we select will of

course be the smile, which is indeed an observable and measurable quantity, with primordial importance w.r.t. to prices, hedges, sensitivities and risk. In essence, we will *replace* some or all of the SinsV model native parameters, (such as vol of vol or correlation) by descriptors of the smile, (static and dynamic, such as curvature or skew) that they generate.

In a sense, this approach can be seen as the “poor man’s stochastic implied volatility model”. But, as we observed previously, the correspondence between the two classes (SinsV and SimpV) is very strong, so that adopting the above modelling strategy consists *de facto* in opting for a tradeoff: what we lose in freedom and precision³¹ of the specification for the SimpV model, we gain in *validity*, as a SinsV model is intrinsically non-arbitrable.

An added benefit of the intuitive re-parametrisation is that it reduces the complexity of “extended” Vega computations. These usually consist in computing the sensitivity of the smile³² w.r.t. to the model parameters, and then inverting (when possible) the Jacobian matrix, in order to evaluate the Vega/smile risk contained by exotic products. With the new representation, the process is simplified because this high-dimensional matrix is reduced or even diagonal. Indeed, the smile in question *is* now (a subset of) the parameter set.

Turning to practicalities, we start by characterising the static shape of the target smile in terms of several of its first IATM differentials: typically the ATM level, skew, curvature and slope for short expiry options. Alternatively, we can express or approximate these quantities with simple options, such as ATM Straddles, Strangles, Call Spreads, Butterflies, Maturity Spreads, etc. In order to compute the native parameters, we can either use the results of the Recovery Theorem 2.1 [p. 55] or invert the formulae of the First Layer Theorem 2.2 [p. 62].

In its principle, the intuitive approach belongs firmly to the inverse problem family; however we will tend to bring the results *via* the direct formulas of Theorem 2.2. The reason is that we demonstrate the method here by focusing on the *static* properties of the smile, which are the easiest to measure and comprehend. Obviously the method can also incorporate the dynamic coefficients: $\tilde{b}$, $\tilde{v}$, $\tilde{v}'_y$ and $\tilde{n}$ for the first layer, all taken in $(t, 0, 0)$. Also, there is a mathematical price to pay, a structural drawback to staying purely with static quantities. Indeed, we have seen with Corollary 2.7 that a bijectivity existed between the $\tilde{\Sigma}^-(2,0)$ and $\sigma_t^-(2,0)$ groups. We have also mentioned in Sect. 2.4.2 that assuming bi-dimensionality and some regularity assumptions, a full bijection could be expected between, on one hand, some groups of instantaneous coefficients $(a_{i,t})$, and on the other hand, collections of IATM differentials. The thing is that those collections of differentials include both static and dynamic quantities. Granted, we have seen that some of these are interchangeable, since within the first layer for instance, we have the following equivalence pairs: $\tilde{v}(\star)$ vs $\tilde{\Sigma}'_y(\star)$, $\tilde{v}'_y(\star)$ vs $\tilde{\Sigma}''_{yy}(\star)$ and $\tilde{b}(\star)$ vs $\tilde{\Sigma}'_\theta(\star)$. But recall also that the dynamic coefficients relate either to the drift or to the endogenous driver. Hence $\vec{\tilde{n}}(\star)$ cannot be swapped, so

³¹ The method is only asymptotic, after all.

³² The surface is usually represented either by a selection of points, or by a collection of parameters.

that removing it from the list of “intuitive” quantities will lose us some information, which we will have to compensate by making supplementary assumptions.

2.5.3.2 Intuitive Models: Examples

First, let us note that we have recently encountered such (potential) intuitive models. Indeed, Corollary 2.12 [p. 101] presents, respectively for the LDD and CEV models, the two native parameters as functions of the IATM level and skew. When limited to their pure local volatility versions, these two quantities define the models entirely. While in their extended (EDCEV) versions, only the perturbation (2.95) remains to be specified. In that case it is possible to use a mixed representation, with both native and implied (intuitive) parameters.

We now provide two distinct examples: note that these are purely “toy” models, or demonstrative instruments. As such we are not looking for high precision in a real-life model: this will be the object of Chap. 4.

The following example does not make use of a “real” SinsV model, but instead uses the generic formulation, the “cast” presented in Sect. 2.1.3.1. It demonstrates how easily we can generate desired surfaces, using sparse, low-dimensional specifications. In this case we have chosen to illustrate with an unusual, probably unheard of, *concave* smile shape. Indeed, result (2.53) implies that the smile generated by a SInsV model can theoretically exhibit negative convexity, at least in the vicinity of the origin $(t, 0, 0)$. In order to verify the validity of this prediction, we use the following toy model, which does not make use of any exogenous driver.

We have also seen that when using the asymptotic methodology, the SDE system describing the model is only accounted for *via* the values at the initial point. As such, this example underlines not only the efficiency but also the limitations of the asymptotic approach.

Example 2.4 (Endogenous SV model generating a concave smile) Assume the following stochastic instantaneous volatility model.

$$\frac{dS_t}{S_t} = \sigma_t dW_t \quad \text{with} \quad d\sigma_t = a_{2,t} dW_t \quad \text{and} \quad da_{2,t} = a_{22} dW_t,$$

where we choose the following initial values and parameter:

$$S_0 = 100 \quad \sigma_0 = 0.2 \quad a_{2,0} = 0 \quad a_{22} = -1.$$

On a short-term basis, there is nothing drastically unrealistic with these dynamics. The level of volatility σ_t is consistent with usual market conditions, the vol of vol a_2 is null but very unstable and negatively correlated with the underlying: all in all, these parameters could economically correspond to a “tipping point” in the market. This configuration typically occurs just before an anticipated and important announcement

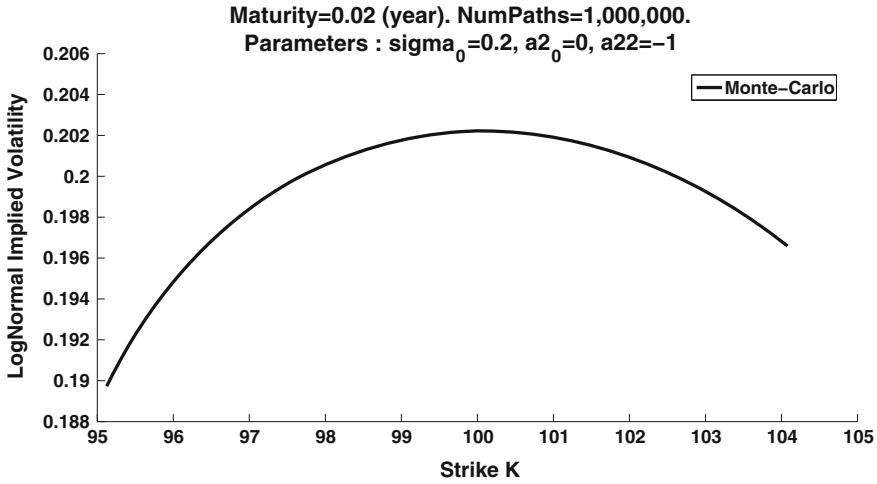


Fig. 2.3 LogNormal implied volatility: Monte-Carlo simulation with Euler scheme and $\Delta t = 10^{-3}$

which could swing either way: a central bank meeting or the release of a major economic indicator.

However, the resulting theoretical short-term smile is very unusual. Indeed, Theorem 2.2, in particular (2.58) and (2.59), predicts no skew and an IATM negative curvature of

$$\Sigma''_{KK}(0, S_t, S_t, 0) = \frac{1}{3} S_t^{-2} \sigma_t^{-2} a_{22,0} = \frac{1}{3} 100^{-2} 0.2^{-2} = 8.3310^{-4}$$

which, using a Maclaurin expansion for a strike of 95, represents a drop of $\approx 1\%$ Lognormal IV.

The actual smile is obtained by Monte-Carlo simulation, with a million paths, for a maturity of a week ($2 \cdot 10^{-2}$ year) and computed for strikes ranging 5 either side of the money: it is shown in Fig. 2.3.

Besides the possible economic interpretations, it is clear that the model served its purpose, as the Monte-Carlo simulation is in accordance with the theoretical prediction.³³

Note that with such a high number of paths the standard error is negligible and therefore was not represented. For readers wishing to reproduce the test, note the issue of boundary conditions and the following identity:

$$a_{2,t} = a_{22} W_t \text{ hence } d\sigma_t = a_{22} W_t dW_t \text{ thus } \sigma_t = \frac{1}{2} a_{22} (W_t^2 - t) .$$

³³ Indeed the level, skew and curvature are conform: check the 95 strike.

Of course, there exists an infinity of SinsV models that will provide the same cast and therefore the same IATM differentials, at least within Layer 1. This is due to the asymptotic nature of the method, which brings two restrictions:

- the processes/coefficients $a_{2,t}$, $a_{3,t}$, etc. are only considered at current time t ;
- any IATM differential will only depend on a finite number of those coefficients.

The first restriction implies that, for instance, the method will be blind to a full time-dependency such as $\{\sigma(s)\}_{s \geq t}$: it will only consider the values $\partial_{s_i}^{(i)} \sigma(s)|_{s=t}$ until a finite differentiation order. The second restriction means, for instance, that if we were interested in $\tilde{\Sigma}_{yy}''(t, 0, 0)$, then that curvature would be unaffected by an over-specification *in depth* of the dynamics of a_{22} . That specification, however, might lead to a significantly different model (in terms of dynamics, long-term marginal distribution, etc).

As a first illustration of these restrictions, it is of course possible to make the model of Example 2.4 even more realistic, especially for longer-term dynamics. We can, for instance, make the volatilities mean-reverting or Lognormal, thus avoiding some of the boundary issues. But all configurations providing the same instantaneous “cast” will generate very similar smiles, at least for such a short, one-week maturity. Accordingly, adding mean-reversion to volatility and vol of vol as per

$$d\sigma_t = \kappa (\theta - \sigma_t) dt + a_{2,t} dW_t \quad da_{2,t} = \kappa_2 (\theta_2 - a_{2,t}) dt + a_{22} dW_t$$

would not modify the IATM level, skew, and curvature. This is because, although this specification creates a (somewhat unusual) multi-scale SV model, it effectively provides an a_1 (and a_{21}) term, which in Layer 1 only affects the slope $\tilde{\Sigma}_\theta'(t, 0, 0)$.

Similarly, and as a second illustration, since the toy model is purely endogenous we can obtain the same instantaneous cast with a pure local volatility model. As before, in order to match the current values of σ_t , $a_{2,t}$ and $a_{22,t}$ quoted above, it suffices³⁴ to specify a concave f as per

$$\frac{dS_t}{S_t} = f(S_t) dW_t \quad \text{with} \quad f(S_t) = 0.2 \quad f'(S_t) = 0 \quad f''(S_t) = -2.510^{-3}.$$

Clearly these conditions constrain very little the *global* behaviour of $f(\cdot)$ as a function, which implies through Dupire that all sorts of smiles could be generated when far from the immediate money. Assuming f is smooth and analytical, the only way through Dupire to extend our control over the smile would be to impose further IATM differentials $f^{(i)}(S_t)$. In our asymptotic framework, this would correspond to obtaining more IATM differentials *via* deeper coefficients $a_{i,t}$: this will be the object of Sects. 3.1 and 3.2.

Let us now illustrate the intuitive re-parametrisation with another toy model:

Example 2.5 (Lognormal model with Normal, correlated stochastic volatility) Let us consider the following bi-dimensional model:

³⁴ The simple proof is left to the reader, see Lemma 2.4, p. 90.

$$\frac{dS_t}{S_t} = \sigma_t dW_t \quad \text{with} \quad d\sigma_t = \varepsilon dB_t \quad \text{and} \quad \langle dW_t, dB_t \rangle = \rho dt \quad (2.99)$$

with $\varepsilon \geq 0$ some constant. The cast is easily deduced from the original model *via*

$$a_2 = \rho\varepsilon \quad \text{and} \quad a_3 = \sqrt{1 - \rho^2}\varepsilon,$$

which justifies that both initial coefficients $a_{2,t}$ and $a_{3,t}$ should be positive.

Let us now assume that the agent is interested purely in the level, skew and curvature of the smile, and that he/she prefers expressing these differentials w.r.t. strike K , rather than log-moneyness y .³⁵ Also, the agent chooses to ignore the slope Σ'_T altogether, which justifies the lack of a drift (e.g. mean-reversion) in the volatility dynamics. The exposed model parameters are now Σ , Σ'_K and Σ''_{KK} (all taken IATM).

Then applying Corollary 2.5, we get the IATM differentials:

$$\Sigma'_K = \frac{a_2}{2S_t\sigma_t} \quad \text{and} \quad \Sigma''_{KK} = \frac{1}{S_t^2} \left[\frac{1}{\sigma_t} \left[-\frac{a_2}{2} \right] + \frac{1}{\sigma_t^3} \left[\frac{a_3^2}{3} - \frac{a_2^2}{2} \right] \right],$$

which we can invert, to express the instantaneous coefficients $a_{2,t}$ and $a_{3,t}$ as

$$a_2 = 2S_t\Sigma\Sigma'_K \quad \text{and} \quad a_3 = \sqrt{3} \left[S_t^2\Sigma^3\Sigma''_{KK} + S_t\Sigma^3\Sigma'_K + 2S_t^2\Sigma^2\Sigma_K'^2 \right]^{\frac{1}{2}}.$$

Note that we need to check the existence of a solution (i.e. that the quantity under the square root is positive) for the given set of IATM differentials. Note also that we could equivalently use the inverse results (2.44) and (2.50) (see pp. 56 and 60), providing they were first converted to K -differentials. Finally, we can use the cast to re-express the native model parameters as implied quantities with

$$\varepsilon = \sqrt{a_2^2 + a_3^2} \quad \text{and} \quad \rho = \frac{a_2}{\sqrt{a_2^2 + a_3^2}}.$$

Although our toy model (2.99) can be compared to Heston's [2], the dynamics of its volatility are Normal (as opposed to C.I.R.) and therefore no (semi-) closed form is *a priori* available. However, note that, were such a closed-form available for the price or even for the implied volatility, the intuitive approach would still be pertinent.

As mentioned above, the inversion technique is by no means always possible. In the current case, having excluded the exogenous dynamic coefficient $\vec{n}(\star)$ from the “intuitive” quantities, we knew that (in the general case) we only had access to the *modulus* of $\vec{a}_{3,t}$. Therefore we had to rely on a bi-dimensional setup (Z_t is here scalar) along with the given, native constraint that $a_{3,t}$ was positive.

³⁵ Recall that with static quantities the absolute/sliding distinction is moot.

This illustrates the fact that, even with valid inputs (the IATM differentials), the existence and/or uniqueness of a solution might have to be enforced by applying supplementary (and possibly arbitrary) constraints. Formally, we first have to ensure that a bijection exists between the collection of IATM differentials that we take as input, and a group of instantaneous coefficients ($a_{i,t}$). This is where further assumptions or constraints might have to be applied, for instance if only shape differentials come as input, in order to establish the bijection. Then we have to ascertain that a bijective relationship also exists between that group of coefficients and the collection of native parameters that we wish to replace.

In practice this process rarely represents an issue, if only because the number of parameters and the order of differentiation are usually limited (typically within the first layer). This inversion technique will be illustrated in detail within Chap. 4, for more complex models such as SABR and FL-SV. We will see that it can also be a quick and powerful tool for the initial calibration, as it provides a close initial guess.

Note also that in this example it seems more appropriate to invert the “direct” results of Corollary 2.5 than to use the “inverse” formulas of the Recovery Theorem 2.1.

In our view, the fact that the asymptotic formulas lose pertinence and precision when far from the IATM point³⁶ does not hamper the relevance of the “intuitive modelling”. The practitioner should not expect to parametrise the model by specifying *ex ante* and with precision any long-term smile. What he/she controls is the very short-term smile, which is certainly more natural than, for instance, gauging the magnitude of the vovol in Heston’s model.

Note also that, as with the EDCEV model, a mixed representation using both implicit and native parameters is certainly conceivable. With model (2.99) we could alternatively specify

- either the IATM skew Σ'_K and the volatility of volatility ε ; or
- the IATM curvature Σ''_{KK} and the correlation ρ .

But as always, any such under-specification or excessive degree of freedom would be seen as an asset for modelling, but also as a liability for the calibration process.

2.5.3.3 Intuitive Models: Practical Usage

So far we have mentioned the principles of the approach, and broached the subject by means of a few simple examples. Beyond the theoretical aspects and all the choices that they offer, we have not discussed how to fit the methodology to market realities, and in particular which smile descriptors *could* or *should* in practice be taken as input. Nor have we offered a typical way of resolving the possible inversion issues that might arise from that choice: these are the goals of the current section. In order to remain as generic as possible, we do not specify an actual SInsV model, but instead we stay at the cast level. Note, however, that few of our following assumptions can

³⁶ See the discussion within Sect. 2.5.1.1, p. 77.

be seen as universal: those are just fairly common features of advanced markets, at the time of writing. We also make choices that are subjective: the point is to walk the reader through the thought process and provide general directions, nothing more definitive.

We wish to specify *ex ante* the most relevant features of the smile, both static and dynamic, through a stochastic instantaneous volatility model. In a trading context, “relevant” is often synonymous with “liquid” and in most single-underlying markets, the most traded assets will be the following, in order of frequency:

1. the underlying;
2. the short-dated ATM volatility level (Calls, Puts, Straddles);
3. the short-dated ATM skew (Binaries, Call spreads, Risk Reversals, Collars);
4. the short-dated ATM curvature (Butterflies, Strangles);
5. the short-dated ATM slope, providing enough liquidity in Calendar Spreads.

All in all, these products need to be matched at any given time, which gives us five *static* targets.

Although our presentation and methodology is both probabilistic and asymptotic, it is not incompatible with some statistical approaches. For practitioners, the modelling and calibration phases are usually torn between the implied and realised arguments: statistical arbitrage and hedging of exotics can seem at odds. Our results allow us to bridge part of that gap and combine both approaches, because they structurally link the statics and the dynamics.

Notwithstanding some dedicated, implied products such as volatility derivatives, statistics in our context means inference of processes, which in turn requires good-quality time series. These must be observable, financial quantities published at a sufficiently high frequency, which limits us therefore to the five static quantities quoted above. Incorporating the numeraire and risk premia defining the measure, it is possible to infer historically their dynamics, i.e. the drifts, volatilities and correlations. We already know, however, that in a SImpV context the drifts are redundant, and we will as usual consider vectorial volatilities, which incorporate the correlations.

When it comes to inferring those volatilities, the endogenous driver W_t should be extracted from the underlying's time series, which is the most liquid. As for exogenous components, the first hurdle is to decide on a dimension for the exogenous driver $\vec{Z}_t$. Given the list of static targets, the maximum useful dimension is 4; the actual dimension will usually be chosen after a Principal Component Analysis of those time series. In most markets though, only the short-dated *level*, and possibly the skew, present a liquidity comparable to the underlying's, and even then it comes at a lower frequency. Therefore all inference of exogenous volatilities will be less precise than endogenous ones. In our view, this statistical feature is structural and justifies not only the use of a single-dimensional Z_t (in this case), but also favouring endogenous volatility components over exogenous ones.

In accordance with these observations, and in view of our established asymptotic results, we will therefore limit ourselves to specifying:

6. the (scalar) volatility of the underlying;
7. the volatility (endog. *and* exog. components) of the short-dated ATM level;
8. the endogenous volatility component of the short-dated ATM skew.

In total, that brings us to nine targets in order to describe the SImpV-*via*-SInsV model. Obviously, the underlying is matched by construction, while σ_t and $\tilde{\Sigma}(\star)$ are redundant. Furthermore, Recovery result (2.37) [p. 54] shows that when specifying the IV statics and dynamics, arbitrage constraints preclude the dissociation of the IATM level $\tilde{\Sigma}(\star)$ and skew $\tilde{\Sigma}'_y(\star)$ from the IATM endogenous volatility $\tilde{v}(\star)$. Similarly, result (2.38) establishes that $\tilde{v}'_y(\star)$ is entirely determined by $\tilde{\Sigma}'_y(\star)$, $\tilde{\Sigma}''_{yy}(\star)$ and $\tilde{n}(\star)$. Together, these equations take us down to only four targets, but also with four instantaneous coefficients: $a_{1,t}$, $a_{2,t}$, $a_{3,t}$ and $a_{22,t}$.

As occurred previously with Example 2.5, the associated inverse problem can then be made well-posed, with the proviso that $a_{3,t}$ be taken scalar and constrained to a given sign³⁷: it suffices to invoke (2.50) (see Corollary 2.4, p. 60).

So in principle we can infer a cast for the SInsV model; how this cast will determine the native parameters is an *ad hoc* issue, dependent on the chosen SInsV class, and might require the introduction of further constraints. More generally, the whole process depends on the collection of IATM differentials that was chosen, along with the parametrisation of the SInsV model. The problem can easily be ill-posed, for instance if we have more parameters than inputs. This kind of situation, on the other hand, is an appropriate context to introduce a more advanced usage of the inverse methodology, namely the *dynamics-influenced calibration*. Let us consider, for instance, the pure (re-)calibration issue of a given SInsV model. On one hand, assuming the underlying's path has been observed, the native parameters themselves can also be inferred (for instance, using likelihood estimators). On the other hand, dependent on which static and dynamic smile descriptors are available, the asymptotic methodology described above provides at least some of the $a_{i,t}$ coefficients, which determine or constrain the native parameters. We can then combine (hedging) or confront (arbitrage) both methods within a multi-objective optimisation process. How well the solution fits each objective is a measure of their compatibility, and therefore provides valuable information to both the arbitrageur and the hedger. The interest for statistical arbitrage is obvious. As for hedging, let us recall that a better dynamic stability of the parameters should also narrow down the tracking error and its variability.

In essence the principle is not new, but simply extended and accelerated by the asymptotic results. Extended because we can now throw in the *dynamics* of the smile, as opposed to only its shape. Accelerated because even the static calibration process is made easier and generic. All in all, we gather these variants of the calibration process under the denomination of *dynamics calibration*, which is strongly linked with but mitigates the process of *re-calibration*.

³⁷ Usually positive as in (2.99).

2.6 Conclusion and Overture

We consider that one of the main contributions of this study (summarised in Fig. 2.4 [p. 113]) is to make explicit the high degree of equivalence between the SInsV and SImpV classes. It is by now apparent that most practical applications will start with

Symbols: $(\circ) \triangleq (t, y, \theta)$ $(\bullet) \triangleq (t, y, 0)$ $(\star) \triangleq (t, 0, 0)$ $(\alpha) \triangleq (t, S_t, K, T)$

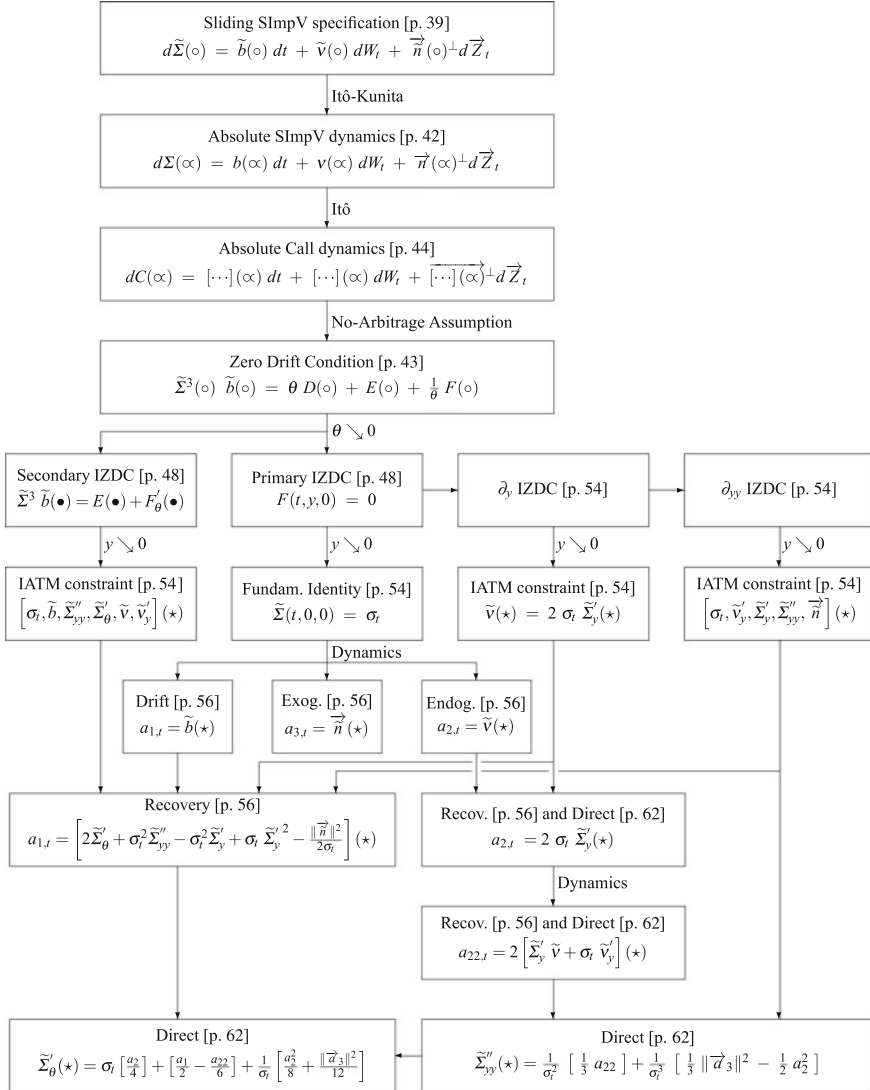


Fig. 2.4 Summary sketch of proofs within this chapter

SInsV models. In these cases the SImpV class becomes a kind of artefact, not destined to be used *per se* but simply to approximate the generated smile shape and dynamics. However, we would not discourage a stochastic implied volatility modelling approach. As mentioned earlier, this model class is rich and potent, although not without implementation difficulties. Overall we see it as one of the future solutions to deal with liquid option markets, but nonetheless we will tend to focus on the direct problem thereafter.

It is also clear that the subject is rich and deserves to be extended in several directions, if at all possible. A first interrogation relates to higher differentiation orders, and the possibility of a generic and programmable approach. In particular, can we compute some of the other meaningful differentials that describe the smile shape (and dynamics)? Another question is the impact of using a multi-dimensional framework, and the way it affects both the direct and inverse problems. Also, as discussed in Sect. 2.1.2.4 [p. 32], is it possible to extend the framework to term structures, and how does it affect the structure of the problem?

References

1. Lewis, A.L.: Option Valuation Under Stochastic Volatility. Finance Press, California (2000)
2. Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Finance Stud.* **6**(2), 327–343 (1993)
3. Black, F., Scholes, M.: The pricing of options and corporate liabilities. *J. Polit. Econ.* **3**(81), 637–654 (1973)
4. El Karoui, N., Jeanblanc-Picqu, M., Shreve, S.E.: Robustness of the black and scholes formula. *Math. Finance* **8**(2), 93–126 (1998)
5. Brace, A., Gatarek, D., Musiela, M.: The market model of interest rates dynamics. *Math. Finance* **7**(2), 127–155 (1997)
6. Schoenmakers, J.: Robust Libor Modelling and Pricing of Derivative Products. Financial Mathematics. Chapman & Hall, Boca Raton (2005)
7. Hafner, R.: Stochastic Implied Volatility. Number 545 in Lecture Notes in Economics and Mathematical Systems. Springer, Berlin (2004)
8. Gatheral, J.: Developments in Volatility Derivatives Modeling. Presentation at the Global Derivatives Trading and Risk Management. Merrill Lynch (2007)
9. Gatheral, J.: Further Developments in Volatility Derivatives Modeling. Presentation at the Global Derivatives Trading and Risk Management. Merrill Lynch (2008)
10. Derman, E.: Regimes of Volatility: Some Observations on the Variation of SP500 Implied Volatilities. Quantitative Strategies Research Notes, Goldman Sachs (1999)
11. Hagan, P.S., Kumar, D., Lesniewski, A.S., Woodward, D.E.: Managing smile risk. *Wilmott Mag.* **1**(1), 84–108 (2002)
12. Durrleman, V.: From implied to spot volatilities. *Finance Stochast.* **14**(2), 157–177 (2010)
13. Alexander, C., Nogueira, L.M.: Stochastic Local Volatility. In: Second IASTED International Conference, pp. 136–141 (2004)
14. Cont, R., da Fonseca, J.: Dynamics of implied volatility surfaces. *Quant. Finance* **2**, 45–60 (2002)
15. Fengler, M.R.: Semiparametric Modeling of Implied Volatility. Lecture Notes, Springer Finance (2005)
16. Ren, Y., Madan, D., Qian, M.: Calibrating and pricing with embedded local volatility models. *Risk Mag. Lond.* **20**(9), 138 (2007)

17. Heath, D., Jarrow, R.A., Morton, A.: Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. *Econometrica* **60**, 77–105 (1992)
18. Schnbucher, P.J.: A market model for stochastic implied volatility. *Philos. Trans. R. Soc.* **357**(1758), 2071–2092 (1999)
19. Ledoit, O., Santa Clara, P.: Relative Pricing of Options with Stochastic Volatility. Technical Report, UCLA (1998)
20. Brace, A., Goldys, B., Klebaner, F., Womersley, R.: Market Model of Stochastic Implied Volatility with Application to the BGM Model. Technical Report, University of New South Wales (2001)
21. Bergomi, L.: Smile dynamics. *Risk* **17**(9), 117–123 (2004)
22. Bergomi, L.: Smile dynamics II. *Risk* **18**, 67–73 (2005)
23. Bergomi, L.: Smile dynamics III. *Risk* **21**, 90–96 (2008)
24. Bergomi, L.: Smile dynamics IV. *Risk* 94–100 (2009)
25. Rousseau, N.: How to Keep the Smile? Dynamic Vega Hedges and Volatility Derivatives. Ph.D. Thesis, Université de Nice Sophia Antipolis (2007)
26. Hull, J., White, A.: Pricing interest-rate-derivatives securities. *Rev. Finance Stud.* **3**(4), 573–592 (1990)
27. Dupire, B.: Pricing with a smile. *Risk* **7**, 18–20 (1994)
28. Lee, R.W.: Implied volatility: statics, dynamics, and probabilistic interpretation. In: *Recent Advances in Applied Probability*. Springer, New York (2004)
29. Avellaneda, M., Lipkin, M.D.: A market-induced mechanism for stock pinning. *Quant. Finance* **3**, 417–425 (2003)
30. Beresticki, H., Busca, J., Florent, I.: Asymptotics and calibration of local volatility models. *Quant. Finance* **2**, 61–69 (2002)
31. Romano, M., Touzi, N.: Contingent claims and market completeness in a stochastic volatility model. *Math. Finance* **7**(4), 399–410 (1997)
32. Dragulescu, A.A., Yakovenko, V.M.: Probability distribution of returns in the Heston model with stochastic volatility. *Quant. Finance* **2**, 443–453 (2002)
33. Lee, R.W.: The moment formula for implied volatility at extreme strikes. *Math. Finance* **14**(3), 469–480 (2004)
34. Benaïm, S., Friz, P.: Regular variation and smile asymptotics. *Math. Finance* **19**, 1–12 (2009)
35. Nicolay, D.: Volatility Dynamics. Ph.D. Thesis, Ecole Polytechnique (2011)
36. Kunita, H.: *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge (1990)
37. Dupire, B.: Pricing and Hedging with Smiles. Working paper, Paribas Capital Markets (1993)
38. Gatheral, J.: *The Volatility Surface: A Practitioner's Guide*. Wiley, New York (2006)
39. Gatheral, J.: Modeling the Implied Volatility Surface. Presentation at the Global Derivatives and Risk Management. Merrill Lynch (2003)
40. Derman, E., Kani, I., Zou, J.Z.: The Local Volatility Surface: Unlocking the Information in Index Option Prices. Quantitative Strategies Research Notes, Goldman Sachs (1995)
41. Hagan, P.S., Woodward, D.E.: Equivalent black volatilities. *Appl. Math. Finance* **6**, 147–157 (1999)
42. Hull, J., White, A.: The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281–300 (1987)
43. Renault, E., Touzi, N.: Option hedging and implied volatilities in a stochastic volatility model. *Math. Finance* **6**, 279–302 (1996)
44. Brigo, D., Mercurio, F., Rapisarda, F.: Lognormal-mixture dynamics and calibration to market volatility smiles. *Int. J. Theor. Appl. Finance* **5**, 427–446 (2002)
45. Alexander, C.: Normal mixture diffusion with uncertain volatility: modelling short- and long-term smile effects. *J. Banking Finance* **28**, 2957–2980 (2004)
46. Piterbarg, V.V.: Mixture of Models: A Simple Recipe for a ... Hangover? pp. 72–77, *Wilmott Magazin* (2005)
47. Fournie, E., Lebuchoux, J., Touzi, N.: Small noise expansion and importance sampling. *Asymptotic Anal.* **14**(4), 361–376 (1997)

48. Henrotte, P.: The case for time homogeneity. *Wilmott Magazine*, pp. 71–75 (2006)
49. Nocedal, J., Wright, S.J.: Numerical optimization. In: Springer Series in Operations Research, 2nd edn. Springer, New York (2006)
50. Andersen, L., Brotherton-Ratcliffe, R.: Extended Libor Market Models with Stochastic Volatility. Technical Report, Bank of America (2001)
51. Piterbarg, V.V.: Stochastic volatility model with time-dependent skew. *Appl. Math. Finance* **12**, 147–185 (2005)
52. Svoboda-Greenwood, S.: The Displaced Diffusion as an Approximation of the CEV. Technical Report, Oxford University, Oxford (2007)
53. Andersen, L., Andreasen, J.: Volatility skews and extension of the Libor market model. *Appl. Math. Finance* **7**, 1–32 (2000)
54. Lee, R.W., Wang, D.: Displaced lognormal volatility skews: analysis and applications to stochastic volatility simulations. *Ann. Finance* **8**, 159–181 (2012)

Asymptotic Chaos Expansions in Finance

Theory and Practice

Nicolay, D.

2014, XXII, 491 p. 34 illus., 26 illus. in color., Softcover

ISBN: 978-1-4471-6505-7