# Chapter 15

# CMS and Static Replication

*A swap rate paid on its own annuity needs no model. Paid on any other date it needs a correction, and the obvious way to find that correction — calibrate a term structure model and ask it — is both more work and less honest than the alternative. This chapter derives the alternative: express the payoff as a portfolio of swaptions held to expiry, and read the answer off one expiry's smile. No short rate model, no term structure of volatility, no dynamics. It is chapter 9's replication argument moved into the annuity measure, and it prices constant maturity swaps and cash settled swaptions the way the market actually prices them.*

## 15.1 A Rate Paid on the Wrong Date

Chapter 8 established the one fact this chapter leans on. The annuity is a positive price, so it is a numeraire, and the swap rate is a price divided by it, so

$$S(t) = \mathbb{E}_t^{\mathrm{Ann}}\big[S(T)\big] \tag{15.1}$$

is a martingale in the annuity measure. A swaption pays off on the annuity and (15.1) is why Black's formula applies to it unchanged.

Now change one thing. Instead of paying $S(T) - K$ over the life of a swap, pay it once, in cash, at a single date. The payoff is the same function of the same rate; only the numeraire it settles against has changed. And (15.1) was a statement about a particular numeraire, so it no longer applies: under the measure the new payment belongs to, the swap rate is not a martingale, and its expectation is not today's forward.

**Remark** (The same mistake, three chapters running)**.** This is chapter 8's future-forward basis again, and chapter 16's quanto adjustment in advance. Each time a quantity that is a martingale under one measure is paid under another, and the gap is a convexity correction. What differs here is the remedy. The future-forward basis was computed inside a model because there was nothing else to compute it from. Here there is: the swaption market quotes the whole distribution of $S(T)$, and a distribution is enough.

## 15.2 The Annuity Mapping Function

The first step is to say what the wrong measure costs. Part of that can be said exactly and part of it cannot, and the chapter turns on knowing which is which.

**Definition 15.1** (CMS Linked Products)**.** A CMS (Constant Maturity Swap) linked product is a contract that references swap rates of swaps of a fixed maturity in their cashflows. For example, a CMS linked strip could pay at every 6 months the swap rate of a swap starting then and ending 10 years from then.

Let us consider a CMS caplet referencing a swap rate S, maturing at time T, and with strike K. This product is cash settled at T and hence the price is

$$\mathrm{CMSCap}(0) = P(0,T)\mathbb{E}^{T}[(S(T)-K)^{+}]$$

Comparing with a swaption, whereas the swaption pays the difference between swap rate and the fixed rate over the life of the swap, the CMS cap settles the difference in cash at the option expiry.

We could in principle calibrate a Hull-White model parameters to market prices of bonds, futures, forwards and swaptions and then use the calibrated model to price the CMS caplet. But let us consider an alternative approach.

Consider a CMS caplet that settles at $T_p \geq T$ which allows for a payment delay.

$$\begin{aligned}
\mathrm{CMSCap}_{T_p}(0) =& P(0,T_p)\mathbb{E}^{T_p}[(S(T)-K)^{+}] \\
=& P(0,T_p)\mathbb{E}^{\mathrm{Ann}}[(S(T)-K)^{+}\frac{P(T,T_p)\mathrm{Ann}(0)}{P(0,T_p)\mathrm{Ann}(T)}] \\
=& \mathrm{Ann}(0)\mathbb{E}^{\mathrm{Ann}}[(S(T)-K)^{+}\frac{P(T,T_p)}{\mathrm{Ann}(T)}] \\
=& \mathrm{Ann}(0)\mathbb{E}^{\mathrm{Ann}}[(S(T)-K)^{+}\mathbb{E}^{\mathrm{Ann}}[\frac{P(T,T_p)}{\mathrm{Ann}(T)}|S(T)=s]]
\end{aligned}$$

Everything to this point is exact. The change of numeraire is chapter 6's theorem and the last line is the tower property, so

$$M(s,T_p) \;=\; \mathbb{E}^{\mathrm{Ann}}\Big[\tfrac{P(T,T_p)}{\mathrm{Ann}(T)}\;\Big|\;S(T)=s\Big],$$

the *annuity mapping function*, is a well-defined object and not an approximation to one. It is also unknown. It is a conditional expectation under the pricing measure, so computing it means knowing the joint law of the ratio $P(T,T_p)/\mathrm{Ann}(T)$ with the swap rate — which is exactly the term structure model this section set out to avoid.

**Assumption 15.2** (Linear terminal swap rate)**.** $M$ is affine in the rate,

$$M(s,T_p) \;=\; a(T_p)\,s + b(T_p). \tag{15.2}$$

This is the modelling content of the chapter and everything after it is again exact and the exactness is easy to read backwards onto (15.2).

$$
\begin{aligned}
M(s, T_p) =& \mathbb{E}^{\mathrm{Ann}}[\frac{P(T, T_p)}{\mathrm{Ann}(T)}|S(T) = s] \\
=& a(T_p)s + b(T_p)
\end{aligned}
$$

Once we have a distribution for $S(T)$ in annuity measure, which can be inferred using swaption replication (looking at market values of swaptions with same expiry and swap end date and various strikes), and the linear functions a and b, CMSCap price is a call on quadratic function of $S(T)$. We do not need a term-structure model like Hull-White to price this.

The conditions below then pin $a$ and $b$. They are no-arbitrage identities that the *true* $M$ satisfies whatever its shape, so they constrain the assumption rather than establish it: they say which affine map to use, not that an affine map is right.

Condition 1. Martingale property.

$$
\mathbb{E}^{\mathrm{Ann}}[\mathbb{E}^{\mathrm{Ann}}[\frac{P(T, T_p)}{\mathrm{Ann}(T)}|S(T) = s]] = \frac{P(0, T_p)}{\mathrm{Ann}(0)}
$$

This implies

$$
\begin{aligned}
\mathbb{E}^{\mathrm{Ann}}[M(S(T), T_p)] =& \frac{P(0, T_p)}{\mathrm{Ann}(0)} \\
\mathbb{E}^{\mathrm{Ann}}[a(T_p)S(T) + b(T_p)] =& \frac{P(0, T_p)}{\mathrm{Ann}(0)} \\
a(T_p)\mathbb{E}^{\mathrm{Ann}}[S(T)] + b(T_p) =& \frac{P(0, T_p)}{\mathrm{Ann}(0)} \\
a(T_p)S(0) + b(T_p) =& \frac{P(0, T_p)}{\mathrm{Ann}(0)} \\
b(T_p) =& \frac{P(0, T_p)}{\mathrm{Ann}(0)} - a(T_p)S(0) \\
M(s, T_p) =& a(T_p)(s - S(0)) + \frac{P(0, T_p)}{\mathrm{Ann}(0)}
\end{aligned}
$$

Condition 2. Definition of annuity in terms of zero coupon bonds.

$$
\Sigma_{i=1}^{N_p} \delta_i^p P(T, \tau_i^p) = \mathrm{Ann}(T)
$$

where fixed leg is the payer leg. This implies

$$
\begin{aligned}
\Sigma_{i=1}^{N_p}\delta_i^p\frac{P(T,\tau_i^p)}{\text{Ann}(T)} =&1\\
\Sigma_{i=1}^{N_p}\delta_i^p\mathbb{E}^{\text{Ann}}[\frac{P(T,\tau_i^p)}{\text{Ann}(T)}|S(T)=s] =&1\\
\Sigma_{i=1}^{N_p}\delta_i^p M(s,\tau_i^p) =&1\\
\Sigma_{i=1}^{N_p}\delta_i^p a(\tau_i^p)(s-S(0))+\Sigma_{i=1}^{N_p}\delta_i^p\frac{P(0,\tau_i^p)}{\text{Ann}(0)} =&1\\
\Sigma_{i=1}^{N_p}\delta_i^p a(\tau_i^p)(s-S(0))+1 =&1\\
\Sigma_{i=1}^{N_p}\delta_i^p a(\tau_i^p)(s-S(0)) =&0\\
\Sigma_{i=1}^{N_p}\delta_i^p a(\tau_i^p) =&0
\end{aligned}
$$

Condition 3. Definition of swap rate in terms of bond prices.

$$
\begin{aligned}
S(T) =&\frac{\Sigma_{i=1}^{N_p}L_i(T,\tau_{i-1}^p,\tau_i^p)\delta^p P(T,\tau_i^p)}{\text{Ann}(T)}\\
=&\frac{1-P(T,\tau_{N_p}^p)}{\text{Ann}(T)}
\end{aligned}
$$

This implies

$$
\begin{aligned}
\mathbb{E}^{\text{Ann}}[S(T)|S(T)=s] =&\mathbb{E}^{\text{Ann}}[\frac{1-P(T,\tau_{N_p}^p)}{\text{Ann}(T)}|S(T)=s]\\
s =&M(s,T)-M(s,\tau_{N_p}^p)\\
=&(a(T)-a(\tau_{N_p}^p))(s-S(0))+\frac{P(0,T)-P(0,\tau_{N_p}^p)}{\text{Ann}(0)}\\
=&(a(T)-a(\tau_{N_p}^p))(s-S(0))+S(0)\\
s-S(0) =&(a(T)-a(\tau_{N_p}^p))(s-S(0))\\
a(T)-a(\tau_{N_p}^p) =&1
\end{aligned}
$$

From the implications of conditions 2 and 3, we can determine the coefficients u and v in the linear map $a(t)=ut+v$, and that would fully specify our linear annuity mapping. We can then compute price of CMS caplet using a terminal swap rate model, without needing a full term-structure model such as Hull-White.

**Remark** (What the assumption buys, and what it costs)**.** Without it, pricing the caplet needs the joint law of the curve with the swap rate, which is a term structure model — calibrated, simulated, and carrying every choice chapter 12 and chapter 13 spent their length measuring. With it, the payoff becomes a function of $S(T)$ alone, and the price then needs only the *marginal* law of $S(T)$ under the annuity measure. That marginal is quoted: it is the swaption smile at that expiry and tenor, read off by chapter 9's theorem. So the assumption is what converts a model problem into a replication problem, and it is why the method exists at all — it prices a CMS from the swaption smile rather than from a model of the curve.

What it costs is harder to state than it looks, because the assumption is better than a first reading suggests and the exposure is somewhere else.

Note first that the mapping can be *derived* rather than assumed, and the chapter's own opening said how: pick a term structure model. In any one-factor model the whole curve at $T$ is a deterministic function of the single state, so the swap rate and the ratio $P(T, T_p)/\mathrm{Ann}(T)$ are both explicit functions of it, the state can be eliminated between them, and no conditioning is required at all — the conditional expectation collapses to a function. In Hull-White the curve is exponential-affine in that state, so the mapping is an explicit ratio of sums of exponentials, and it is certainly not a straight line.

But it is very nearly one. Over a band of two hundred basis points either side of the forward, the exact Hull-White mapping departs from its own best straight line by about a fifth of one per cent of its level.[1] Assumption 15.2 is therefore a good description of the mapping's *shape*, and the worry it invites — that a curved object is being straightened — is not where the money is.

Where the money is, is which straight line. The slope the model gives depends on the mean reversion: at $\kappa$ near zero it is 0.63, and at $\kappa = 0.10$ it is 0.70, eleven per cent higher.[2] That is fifty times the shape error, and it is a fact about a parameter chapter 12 identifies as the largest model risk in a callable book and as invisible in the instruments used to calibrate.

So deriving the mapping from a model does not remove the assumption. It renames it. A free coefficient $a(T_p)$, pinned by the arbitrage conditions up to one number, becomes a coefficient determined by $\kappa$ — and $\kappa$ is not quoted. What is gained is that the remaining freedom now has a name, a plausible range, and a literature; what is not gained is a number the market has told anyone.

Where the curvature does show up is far from the money, and it grows about as fast as the square of the distance: a twentieth of a per cent of the mapping's level at a hundred basis points either side, a per cent and a half at six hundred.[3] That matters because the replication below weights the wings, so the one region where a straight line is a poor fit is the region the method leans on. It remains a smaller effect than the slope.

It is not, incidentally, made worse by a payment delay. The natural guess is that delaying the payment makes the mapping matter more and bends it further; the opposite happens. A delayed bond falls with the rate as well, which partly offsets the annuity falling underneath it, so the slope drops from 0.64 to 0.35 as the delay runs out to two years and the departure from a straight line falls with it.[4]

So the recommendation to price the trade a second way and reserve the difference, which chapter 20 makes generally, needs sharpening here. The second way *is* a term structure model, and the model is where the mapping came from — so the two prices do not differ by method at all. They differ by $\kappa$, and the reserve to carry is a reserve against mean reversion.

---

[1] `measured`.

[2] `measured`.

[3] `measured`.

[4] `measured`.

## 15.3 The Payoff as a Portfolio

Everything above is exact and none of it is yet computable, because $M$ multiplies the payoff inside an expectation whose distribution we have not written down.

**Theorem 15.3** (Static replication)**.** *Let $g$ be twice differentiable and let $F = S(0)$. Then for every $s$,*

$$g(s) = g(F) + g'(F)(s-F) + \int_F^\infty g''(K)\,(s-K)^+\,dK + \int_0^F g''(K)\,(K-s)^+\,dK. \tag{15.3}$$

*Proof.* Taylor's theorem with integral remainder, split at $F$. For $s > F$, $g(s) = g(F) + g'(F)(s - F) + \int_F^s g''(K)(s-K)dK$, and the integrand vanishes for $K > s$, so the upper limit may be taken to infinity with $(s-K)$ replaced by $(s-K)^+$. The case $s < F$ is the same computation downwards. □

Equation (15.3) is an identity in $s$ and not an approximation, so it can be checked at a point rather than only through the prices it produces — and it is, for payoffs of both signs of curvature, at levels either side of the forward.[5]

**Structure** (Replication is a change of basis)**.** Theorem 15.3 is easier to hold on to as a statement about coordinates than as an identity to verify.

Payoffs on a single rate form a vector space. Equation (15.3) exhibits a basis for it — a constant, a forward, and the calls and puts at every strike — and gives the coordinates of an arbitrary payoff in that basis, namely $g''(K)$. Pricing is then linear algebra: apply the pricing functional to each basis element, which the market has already done and published as a smile, and take the same combination.

Chapter 9's Breeden-Litzenberger is the dual statement in the same basis. There the coordinates of the *density* were extracted from prices; here the coordinates of the *payoff* are extracted from the payoff. The two are the two halves of chapter 3's pairing, and the reason both are model-free is that neither ever asks how the underlying moves — a basis decomposition is not a dynamical statement.

**Remark** (What it says)**.** Any European payoff on the swap rate is a portfolio: some cash, a forward, and a continuum of swaptions with weights $g''(K)$. The portfolio is bought once and held — nothing is rebalanced, no hedge ratio is computed, no process is assumed. This is chapter 9's Breeden-Litzenberger read in the other direction: there, option prices gave the density; here, the density is never formed and the option prices are used directly.

Taking expectations in the annuity measure, where $\mathbb{E}^{\mathrm{Ann}}[S(T) - F] = 0$ by (15.1),

$$\mathbb{E}^{\mathrm{Ann}}[g(S(T))] = g(F) + \int_F^\infty g''(K)\,C(K)\,dK + \int_0^F g''(K)\,P(K)\,dK, \tag{15.4}$$

with $C$ and $P$ the payer and receiver swaption prices per unit annuity — which is to say, the market's own quotes with the annuity divided out.

[5] `checked`, together with `the fact that the expansion point is arbitrary`: expanding about the forward is a convenience that makes the linear term vanish in expectation, not a requirement of the algebra.

**Calculation 15.4** (The adjustment, from the smile)**.** Put the two halves together. With the linear map, $M(s) = a(s-F) + P(0,T_p)/\mathrm{Ann}(0)$, and the CMS swaplet is worth

$$\begin{aligned}\mathrm{Ann}(0)\,\mathbb{E}^{\mathrm{Ann}}\big[S(T)M(S(T))\big] &= \mathrm{Ann}(0)\Big(a\,\mathbb{E}^{\mathrm{Ann}}[S(S-F)] + \tfrac{P(0,T_p)}{\mathrm{Ann}(0)}\,\mathbb{E}^{\mathrm{Ann}}[S]\Big)\\ &= \mathrm{Ann}(0)\,a\,\mathrm{Var}^{\mathrm{Ann}}(S(T)) + P(0,T_p)\,F,\end{aligned}$$

so that

$$\mathbb{E}^{T_p}[S(T)] - S(0) = \frac{\mathrm{Ann}(0)}{P(0,T_p)}\,a(T_p)\,\mathrm{Var}^{\mathrm{Ann}}\big(S(T)\big). \tag{15.5}$$

And the variance is (15.4) with $g(s) = s^2$, whose second derivative is the constant 2:

$$\mathrm{Var}^{\mathrm{Ann}}(S(T)) = 2\left[\int_F^\infty C(K)\,dK + \int_0^F P(K)\,dK\right]. \tag{15.6}$$

**Remark** (Read what is on the right hand side)**.** Nothing in (15.5) or (15.6) is a model output. The annuity and the discount factor are chapter 7's curve. The coefficient $a$ came from three no-arbitrage conditions. The variance is an integral of quoted swaption prices across strikes, with equal weight at every strike — which is the same object a variance swap pays, met here in a rates market instead of an equity one.

So the CMS convexity adjustment is a statement about the swaption smile and nothing else. A desk that has calibrated a Hull-White model to price it has done a great deal of work to obtain an answer that its own screen already contained, and has additionally made the answer depend on the model's parameterisation — a one factor Gaussian model has a particular smile, generally not the market's, and the CMS price inherits it.

**Remark** (Why the whole smile and not the at-the-money)**.** Equation (15.6) weights every strike equally. That is the sentence to remember, because it explains why CMS is a wing product: a payoff that is quadratic in the rate has constant $g''$, so the far strikes contribute as much per unit of strike as the near ones, and they are precisely the strikes with the least liquidity and the widest bid-offer.

It is also why a CMS book is exposed to the volatility of volatility rather than merely to volatility. Raising $\nu$ in the SABR model of chapter 10 lifts the wings while leaving the at-the-money almost unmoved, and (15.6) charges for the wings.

## 15.4 Cash Settled Swaptions

The same defect, in a product that looks like it should not have one.

A cash settled swaption does not deliver a swap. It pays a cash amount computed from the swap rate at expiry using a *formula* for the annuity rather than the annuity itself — typically

$$\mathrm{Ann}_{\mathrm{cash}}(S(T)) = \sum_{i=1}^{N} \frac{\delta}{(1+\delta S(T))^i},$$

a function of the single rate, standing in for a sum of discount factors that are not all determined by it.

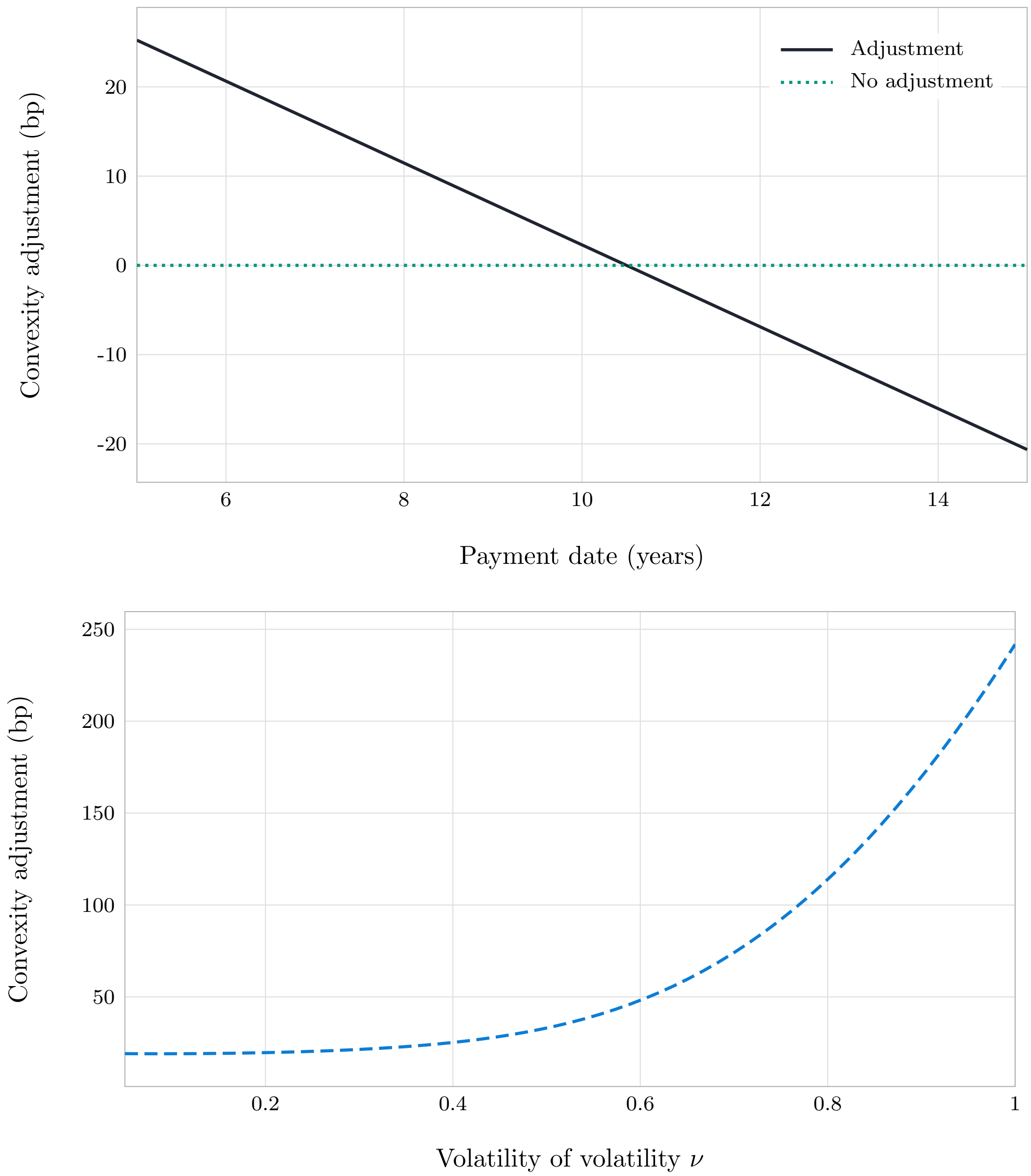


**Figure 15.1**: The CMS convexity adjustment on a five year expiry into a ten year swap, computed by replication over a SABR smile. The left panel is the adjustment against the payment date. It is largest when the rate is paid at the start of the swap it references, falls as the payment is delayed, and crosses zero — the coefficient $a(t)$ declines across the swap, so far enough out the timing effect more than cancels the convexity. The right panel is the adjustment against the volatility of volatility at a fixed at-the-money level, which is flat only if one believes the wings do not matter. They contribute equally per unit of strike, by (15.6), and a CMS book is long them.

Drawn by `convexity_adjustment` and `annuity_variance`.

**Remark** (Why this is the same problem)**.** The physically settled swaption pays $\mathrm{Ann}(T)(S(T)-K)^+$ and the annuity in front is the numeraire, so it divides out and Black's formula survives. The cash settled one pays $\mathrm{Ann}_{\mathrm{cash}}(S(T))(S(T)-K)^+$, and $\mathrm{Ann}_{\mathrm{cash}}$ is *not* the numeraire — it is a different function that happens to agree with it on average. The ratio $\mathrm{Ann}_{\mathrm{cash}}(S)/\mathrm{Ann}(T)$ does not divide out, and what is left is again an expectation of the payoff against something correlated with it.

So the two are not the same trade with a settlement convention between them. They differ by a convexity correction of the same species as the CMS one, and it is computed the same way: the payoff is a function of $S(T)$ alone, therefore Theorem 15.3 applies, therefore it is a portfolio of ordinary swaptions and the smile prices it.

**Exercise.** A cash settled payer swaption struck at the forward is worth less than the physically settled one at the same strike. Show this from the fact that $\mathrm{Ann}_{\mathrm{cash}}$ is a decreasing function of $S$ while the payoff $(S-K)^+$ is increasing, so the two are negatively correlated. (Hint: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] + \mathrm{Cov}(X,Y)$, and the covariance of an increasing and a decreasing function of the same variable is negative.) Deduce that the difference grows with volatility, and explain why the two conventions were therefore never interchangeable, even before anyone computed the size of the gap.

## 15.5 What This Chapter Did Not Need

No short rate was modelled. No volatility was made a function of time. No parameters were calibrated to a surface, and nothing was simulated. The inputs were a discount curve, a swaption smile at one expiry, and three conditions that follow from the definitions of an annuity and a swap rate.

**Remark** (When a model is unavoidable)**.** The argument above does not generalise as far as it first appears.

Static replication works when the payoff is a function of *one* rate observed at *one* date. Then the smile at that expiry is the whole of the relevant distribution and there is nothing left to assume. A CMS caplet qualifies. A cash settled swaption qualifies.

A Bermudan swaption does not. Its value depends on the joint behaviour of the rate at several exercise dates, and no collection of European quotes determines a joint distribution — which is chapter 18's point about marginals arriving again. A callable range accrual does not qualify either, nor does anything whose payoff depends on the path. For those a model is genuinely required, and which model is the subject of chapter 20.

## References

- Hagan, P. S. (2003). Convexity conundrums: pricing CMS swaps, caps, and floors. *Wilmott Magazine*, March, 38–44.

- Carr, P., & Madan, D. (2001). Optimal positioning in derivative securities. *Quantitative Finance*, 1(1), 19–37.

- Mercurio, F., & Pallavicini, A. (2006). Smiling at convexity: bridging swaption skews and CMS adjustments. *Risk*, 19(8), 64–69.