Nine to Five

17

We now try to examine our risk measures and their properties in the practical context of daily risk management. We structure this chapter by the most common questions you might face, in reverse order of urgency.

17.1 Why Did My VaR Increase?

This is the most common question you will encounter. It will be raised by the CEO, desk managers, traders, and just about anyone in the risk audience.[1] A VaR blowup means increasing capital requirements, may cause breaches of VaR limits imposed on desks or traders, indicate problems with a hedge, signal a pick-up in market activity, etc. The opposite question—"Why did my VaR decrease?"—is much rarer, albeit just as warranted.

The sheer number of positions and risk factors, i.e., the vast amount of raw data available, combined with the (desirable but difficult-to-segregate) portfolio effect can make it tricky to immediately assess specific reasons for VaR movements. To answer this question and to localize any root causes of a VaR change, we basically try to narrow down this large search space.

But before we delve into our data, we should check the calculation for profane mishaps. Maybe a position source system has failed to deliver its data, and instead of the usual 1,000,000 positions, we have calculated only half of them? Maybe some fallback mechanism for such missing feeds (usually just copying the previous day's corresponding data over) has failed? Maybe a market data system has delivered a clearly wrong market value that escaped the data scrubbing team, causing a huge spike and an increase in the target volatility of a dominant risk factor? Maybe all positions are there and priced alright, but the definition of which positions belongs

---

[1]We like to treat the VaR as negative value to preserve consistency with support measures like the VaR-contribution. We quickly translate "increase" into its opposite and keep mum about it.

Table 17.1
VaR-contribution

Risk factor	Sensitivity	VaR-sensitivity	VaR-contribution
FX-EUR	−12,228	176.45	−2,157,701
FX-RUB	−661	152.43	−100,727
FX-GBP	537	−104.10	−55,889
FX-CNY	−119	182.50	−21,641
IR-EUR-Y30	21,668	−0.61	−13,119

to which portfolio was messed up due to a small typo during yesterday's otherwise minor change to the portfolio definitions? Robust automatic check procedures, detailed and human-readable logs, and warning or error alerts will greatly help account for such adverse technical headwinds.

Once such mundane reasons are ruled out, we can proceed to see whether market changes or changes in our positions cause our suspicious figures. A quick cross-check is to calculate yesterday's positions with today's scenarios, and vice-versa. This is a good indicator of which investigative path to go down next.

On the market or risk factor side, the partial VaRs like VaR^{FX} or VaR^{IR} (and changes or non-changes in some of them) help restrict the search scope to types or subsets of risk factors. We can then determine the most important risk factors there (via the portfolio's absolute sensitivities) and verify whether the corresponding target volatilities have changed.

A more elegant and comprehensive way is to check the risk factors with the most negative VaR-contributions—this measure helpfully already intermingles sensitivities and volatilities in a consistent way. Table 17.1 shows an example report of the risk factors with the most negative VaR-contributions for some portfolio. Comparing such a report with, e.g., the previous day's one quickly reveals the comparative relative impact of the various risk factors and sniffs out the combined effect of changes in volas and/or sensitivities. Or to put it more prosaically: in isolation, a seemingly tiny IR vola increase from 0.0003 to 0.0005 might escape the quick human glance amid apparently larger vola movements; the VaR-contribution gives such changes the appropriate sensitivity-weights and is thus able to denounce a change as influential or less so.

On the position side, we should first be aware of this:

- Adding a position to a portfolio, while usually amplifying the VaR, can also dampen it, e.g., when adding a position that acts as a partial hedge to the remaining portfolio. In the extreme case of adding the portfolio $-\Omega$, we can even reduce the VaR to zero.
- Removing a position from a portfolio, while usually curbing the VaR, can also magnify it, e.g., if we remove one of two deals that hedge each other.[2]
- Even positions that idly remain in a portfolio can affect the VaR in both directions, for their contract terms might trigger changes in their PnL characteristics.

---

[2] As we want to avoid the ambiguous "increase" and "decrease" for the negative VaR, the thesaurus is having a field day.

So we need to go about our analysis accordingly and cannot limit our focus to new positions alone. With that in mind, how to best drill down to a position that moved the VaR?

It is tempting to first examine the positions' very tangible sensitivities. But while feasible, this is often quite tedious. There are, plainly, lots of sensitivities on the position level (10^6 times the average number of risk factors the positions are sensitive to). Many of them offset each other and blur the picture. Unsuspiciously small sensitivities might stem from non-linear positions and just camouflage their severe tail losses. And ostensibly large sensitivities might have a low corresponding volatility and thus a negligible impact.

Individual VaRs for each position often don't offer very clear signals either, as portfolio effects are not accounted for. Hedges, if not filtered out beforehand and dismissed for their zero risk contribution, will show up twice: linear hedges with two identical (potentially large and alarming) individual VaR values, and non-linear ones, even worse, with different individual VaRs. The former obfuscate the picture somewhat more openly (two positions with the initially troubling but exact same VaR of −217,244.63 are, after all, almost certainly hedges to be dismissed in a quick exploratory analysis). The latter can be more baffling, as their individual VaRs can differ arbitrarily. These very same issues also haunt the incremental VaR of positions.

A more useful hint comes from the positions' cES values. Positions with large negative cES values drive the VaR. And as the cES values of (both linear and non-linear) hedges have opposite signs and the same magnitude, hedges can often be reasonably identified and "guessed away," leaving the remaining positions with dominant cES values as analysis candidates. Still, if there are overly many hedges or imperfect ones that do not close each other completely, they will continue to fog any quick and mindless drill-down analysis of a large set of positions.

How to get around this issue of hedges? In an ideal world, positions that hedge each other are tagged, and you can thus filter them out explicitly. If not, you can try to automatically detect hedges by pair-wise comparison of the positions' basic attributes: if you detect two bonds with the same maturities and coupons but with opposite nominals N and $-N$, they act as hedges and can be removed from consideration. Finally, you could attempt to identify hedges by the positions' sensitivities. If a deal's sensitivities $\mathbf{s}^\alpha$ complement another one's $\mathbf{s}^\beta$ via $\mathbf{s}^\alpha \approx -\mathbf{s}^\beta$ suspiciously closely, those two deals likely hedge each other and can often plausibly be ignored in the search for influential positions. But even in a best-case scenario with all hedges explicitly known and accounted for, positions will often, by chance, at least partially act like hedges to parts or all of the remaining portfolio, so the deficiencies of our support measures with regard to offsetting behavior will linger.

A fast impact analysis can be performed via the visualization presented in Chap. 13: a scatter plot of all deals' PnL standard deviation versus their cES. This allows you to detect the largest and most influential positions as asymmetries in the various characteristic parts of the plot, with the fuzzy human eye readily and helpfully filtering out hedges both perfect and imperfect.

17.2   How Will the VaR Change?

When a new position is about to be added to a portfolio, the effect on the VaR is clearly of interest—after all, it might increase the capital requirements or cause limit breaches. For relatively small positions, a fast shortcut to approximate the VaR impact is to use the portfolio's VaR-sensitivities (see Sect. 7.2). For this, we multiply the new position's sensitivities with the portfolio's corresponding VaR-sensitivities and simply add the resulting dollar terms to the portfolio VaR. This approximation is quick and often close enough. (Note: the signs matter, so if a deal is, e.g., sensitive to the 10-year EUR interest rate to the tune of -1000 and the respective VaR-sensitivity is -4, than the approximate impact on the VaR is $+4000$. This is added to the negative VaR—so $|\text{VaR}|$ decreases.)

A more tedious yet truly precise method is to simply recalculate the VaR with the new position included in the portfolio. As we usually already know the existing portfolio's PnL vector, it is sufficient to compute the new position's one and to then add those two vectors before extracting the resulting new VaR. This of course corresponds to the incremental VaR or pre-deal inquiry mentioned in Chap. 11.

A closely related question is the following: what size should a new position have in order to achieve a certain target VaR? This target is often some optimal or capital-minimizing VaR. To obtain it, we just recalculate the VaR repeatedly, each time adding a different multiple $a\alpha$ of a deal α to the portfolio. (We can even calculate the position's PnL vector only once and then just scale it by a before combining it with the portfolio's vector.)

How will the resulting $\text{VaR}[\Omega + a\alpha]$ look under various position sizes or scaling factors a? As we increase a arbitrarily towards ∞, the position $a\alpha$ will at some point become the dominant deal in the portfolio, and the VaR will get ever more negative. The same happens if we decrease a towards $-\infty$, as the deal once again will become dominant at some point. So we expect to see some sort of upside-down parabola whose maximum or least negative value corresponds to the "minimally risky" VaR. Figure 17.1 depicts an example where a weight of about $a = 0.8$ would lead to minimal capital requirements.
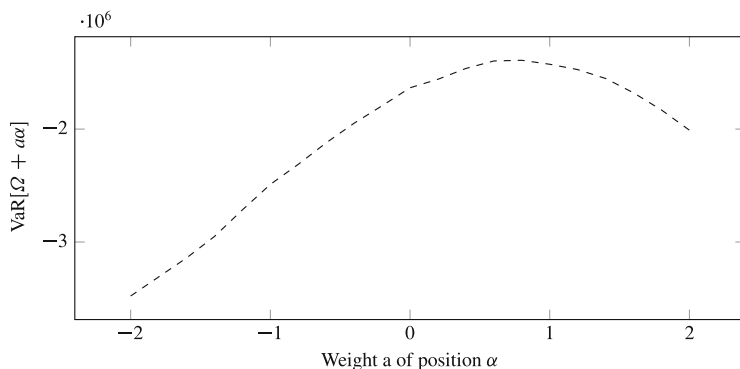

Fig. 17.1 Portfolio VaR under a variable new position

17.3   How Good Is the VaR Model Anyway?

The VaR is—like any summary statistic—a stark abbreviation, in this case of the wealth of information contained in a full PnL vector expressing the two-dimensional concept of risk. As such, the VaR may potentially hide much. It does capture one specific slice or aspect of risk but is probably most useful in a comparative sense (over time or between portfolios) rather than via its raw absolute level. The VaR is not the full picture of risk.

The data used to calculate it is real-world data, prone to errors, missing values, outliers, gaps, etc. The sample size (usually 2 years of data) is small with respect to the large number of risk factors or dimensions. The VaR model's quality much depends on the data fed into it; that data is not always pretty.

On top of that, to make the VaR model react quickly to volatility level changes, it is heavily driven by the last 20 days and their observed returns. This naturally involves a sample error, which is more pronounced for portfolios depending on few risk factors only. If the Monte Carlo modification is applied, an additional noise affects the VaR figure (see Chap. 14). Actual VaR results are fickle.

Nor can we simply math our way out of such issues. Alternative risk measures like the expected shortfall do have some desirable properties, but are still limited by the same basic restrictions. If anything, elaborate models may be too impressive for their own good, lulling the audience into a false sense of certitude and complacency. (More on this in Chap. 13 and the first part of Chap. 9.)

While it is a natural mental crutch to believe that the VaR or a VaR-like measure predicts the future, a better and more prosaic view is that it merely extrapolates some plausible asset behavior from that of the recent past. It assumes that tomorrow will behave like the last few weeks. If this assumption fails, the model's results become void. (For example, a central bank's decision to suddenly devalue a currency might cause a backtesting violation—yet the model couldn't possibly have accounted for it.) This must remain the realm of the underappreciated stress tests (see Chap. 6). The VaR can't foresee the future; it disregards the unrecorded risk.

These reasons should give anyone pause before overestimating the importance or reliability of the VaR. On the other extreme of the pendulum's swing, the VaR should not have to bear the blame for financial calamities—as some would have it in light of financial market crises. Instead, the man relying blindly or solely on it is the culprit, abetted by him who burdens this humble measure with elaborate but difficult-to-vindicate risk quantization duties. The overconfidence at the heart of it is not rarely fueled by mathematically impressive but less-than-transparent model choices.[3]

---

[3]Take, for example, VaR models that must estimate not overnight or 1-day PnLs, but 1- or 3-month PnLs. They may rely on correspondingly large historical returns, but this would make the returns either fewer or older or overlapping. Alternatively, they may try to use daily returns and project them farther into the future, which raises questions of reversion to the mean, among others. The frowned-upon traditional shortcut is to simply scale up the 1-day VaR by the square root of time (in days). Strictly speaking, this is somewhat off or incorrect, but it seems to usually do the trick. A

By construction, the VaR model will behave—within its limitations and its central assumption of "the world stays the same"—as expected and by and large correctly guess tomorrow's market volatility. Still, the model must pass at least some quality checks, and be it only to ensure the absence of programming mishaps. For this purpose, the plain backtesting or more elaborate distribution tests (Chaps. 15 and 16) serve well. Partial VaRs and synthetic marginals (Chap. 11) help plausibilize the model behavior on a more granular level, as the overall, aggregate numbers might hide erratic effects that cancel each other out. Going much further in discussion or deed often seems to be unwarranted—a simple model first and foremost calls for simple workability tests.

17.4   Hmm, How Did the Parameters Materialize?

An ideal model is one devoid of free parameters, one that "just works" without tinkering. John von Neumann once stated: "With four parameters I can fit an elephant"—aptly hinting at the problem of too much modeling leeway.

Now, our VaR model comes pretty close to that ideal. It has few parameters, and all have an immediate, transparent interpretation. We typically use 500 historical returns and rescale them to the recent 20 days' volatility (maybe flooring the target vola at 20% of the overall historical volatility). When using Monte Carlo or injecting kurtosis, a few more setscrews are introduced. If using log-normal interest rate returns, a certain shift is required to be able to account for negative returns. We won't lose much sleep over these settings.

But while these choices are readily made, they can be much more tricky to justify and defend. Especially those parameters without a natural optimality to them, i.e., those not resulting from an accepted calibration procedure, raise doubts. Rightfully probing questions from the regulator ("Why use 20 and not 30 days for local volas?") have no obvious retort. The answer of first resort—"expert opinion"—seems terribly smug; the contrasting alternative—"common sense"—amounts to an insult. Alas, there is often no very satisfactory answer.

There are, however, some avenues in responding that may at least minimize the back-and-forth required to allay such parametrization concerns. One can try, first, to kill the question in the crib and openly shame the parameters from the start as being arbitrarily set: "These parameters have wiggle room—there is no obviously correct or optimal setting. We discussed them and our consensus fell on this particular tuning; we feel that it is reasonable." Your audience, hopefully impressed that you didn't try to bury the issue but to highlight it, might feel less inclined to dig deeper; the hole is already there and plain to see.

A bit more suave it is to rechristen the parameters and declare them to be *meta parameters*—not really within the model scope but somewhat residing ethereally

---

proper long-horizon estimate, on the other hand, has ample leeway in terms of modeling and thus allows for a wide range of results, which doesn't necessarily invite much greater confidence in it.

outside of it. A good example is the choice of the 20-day window for target and local volas. Longer windows cause the VaR to be more stable and to react slower to market changes, while shorter ones do the opposite. This setting is due to the model users' preferences, not to some parameter optimality.[4] In any case, don't ever casually observe that you "tested" various setups (in an feeble attempt to make your point stronger). That will trigger the immediate, checklist-ticking "Can you please send us your test results?" Now you'd actually have to run those tests.

If the "questicide" fails, an answer must reluctantly be produced. Luckily, people since Hammurabi have a fond credulity for all things written. The "I told you so" gets more heft when framed as "This is already written in stone." Try hard to find a reference to a book or paper or institution that has already put forward what you are trying to do.[5] The best reference is NASA, but even an—as of yet illusive—Journal of Chart Analysis might handily absolve you from having to admit that no one ever has come up with your exact same ideas.

Unable to unearth a precedent, you can set up some committee responsible to agree on parameter values, along with periodic meetings, minutes, and stamp-wielding due process. Parameter values conjured up in such a setting are coated in paper-trailed consecration and less likely to be doubted.

If all fails, you are left to prove that other, similar parameter choices just don't make much of a difference. You could, for example, run a parallel calculation of the same model using 30-day instead of the usual 20-day windows. If the VaR time series obviously look very similar, then the model's *parameter sensitivity* is small, and the choice thus inconsequential and unworthy of investigative zeal. This is relatively costly in terms of setup and calculation efforts, but it might well be necessary to quell a topic.

All of the above is better, by the way, than trying futilely to prove some tuning parameter's "optimality"—any utility criterion you maximize will draw attention to its own shaky justification. Instead, as very last resort, you must rely on the gravitas of some PhD-adorned team member, preferably sporting a beard gray and flowing.

Needless to say, this section exists because we got stung. Be prepared to spend time on your parameter defense and reluctant of any model improvement entailing new free parameters (e.g., supposedly superior exponentially weighted volatilities with some—new!—decay parameter). We spent more time explaining parameters we considered obvious (and blatantly were not by our counterparts) than on actually programming the model. It was us who did not press them effectively from the start. The questions on model parameters are comprehensible, warranted, and indeed necessary. Our answers, however, were at times hesitant, piecemeal, or winding. You better best that.

---

[4]Other out-of-scope parameters are *fudge parameters*; they are introduced to circumvent numerical issues. An example is the shift used to make square root processes operate on negative interest rates.

[5]A good example is the RiskMetrics decay parameter setting of 94%. It often goes unscrutinized because it is so commonly used and ingrained in many model rehashes.

17.5   Can You Validate Your Model?

Answering this innocuous question may become formidably time-consuming fast. It is usually posed formally by entities like the regulator or audit, and it stubbornly clings to life in imposed conditions and follow-up requests. The stakeholders' incentives, their familiarity with a model setup, and their available time to dig into the matter vary—causing expectations to diverge:

- It starts with the poor choice of wording: the rather generic "validation" is bound to be interpreted differently by everyone involved. This is compounded by the fact that the actual people representing, for example, the regulator, will change over time, contributing novel and sometimes different views. Regulations or laws seldom offer explicit guidelines or unambiguous criteria.

  Worse still, you can technically *invalidate* a model and prove it to be deficient, but never really completely prove that it works—you might always be just one further test away from uncovering a major weakness.

- Charged with meaning in applied statistics, the word "validation" leads some to immediately focus and hone in on issues like hypothesis testing and distribution assumptions, possibly neglecting that, fundamentally, there is not a lot of statistics going on here. It is sure fun to ponder the use of a battery of statistical tests—more fun probably than stepping back in humility before the challenges in data or operations. Not infrequently have I encountered people that will gladly spend time on the most peripheral statistical aspects while not even contemplating looking at the 2200 risk factor time series—the actual core of it all. That would (obviously) be time-consuming and boring, but there is more to it than that. It is also just not on many people's radar, primed by the perfect data sandbox exercises prevailing in education and academia.

- Vague terms also naturally inspire a longing for formal frameworks and crutches to somehow pin it all down. Expect requests for "validation concepts" (essentially meta documents), for separate "initial model validations" and "continuous model validations," for "actionable" traffic-light score cards, and for explicit triggers and contingency plans (even to hitherto unknown events). While impressive at face value, this whole superstructure is built on the same unchanged, humble ground and more likely to cause sweat than make sense.

  Such a conceptual overhead tends to provide little relief in terms of requirement consistency. Over the years, we have been variously tasked to either provide *more* tests for increased coverage, or *fewer* ones for better readability and accessibility; to perform the *same* tests year in and year out for consistency, or to do *varied* ones to address different model aspects as markets change. These requests—all perfectly worthy and sensible on their own—are hard to reconcile.

What is there to be done? Well, first, avoid the word "validation." Rephrase it as test or, though unsexy and a mouthful, as plausibilization. No harm ever came from this (except to your tongue).

Second, make sure to impart that a VaR model is simple at heart; that the statistical assumptions are trivial; that market volatilities are captured by construction; that political events are out of scope; and that data and its limitations are front and center. You bolster your case if you avoid proprietary model customizations and opaque language or statisticalese. So curb you quants; smite them abacuses—stick to standard methods and familiar terms instead.

With all parties on the same page about these basics, *plausibilize* (see?) your model with the basic backtesting procedures and parameter justifications outlined in Sects. 17.3 and 17.4. Mention the VaR measure's inherent noise (Chap. 14) if you sense concern over minor model result deviations; address worries about distribution tails as needed (Sect. 17.6). Have your p-values ready, but provide them with context: a test might fail for risk-averseness, which is tolerable; and out of many tests, some are actually expected to fail statistically. Above all, visualize the main results—a graph grasped equals trust earned.

As for the overall scope and shape of the results, it would be fantastic to get away with providing a succinct, small subset of varying custom tests as a readable, crisp model quality characterization. Alas, I fear that this is not workable in practice. The regulator is very rightfully bound to distrust a bank's motives—what is to prevent one from running ten tests and reporting only the three best results? Test types should probably remain largely unchanged. As for sheer test counts, using fewer tests may be just as expressive and much more lightweight and accessible—but only as long as everyone involved agrees to this virtue of brevity. This is untenable, as different and changing personnel will require just another type of test, one more metric, or a higher granularity of sub-portfolios. Stemming this tide of requirement creep is hard, and it is probably better to instead just swim along with it by providing a large set of automatically performed tests. Such cover-our-ass results, however unwieldy and ponderous, are difficult to argue against. The world of VaR model plausibilization is one of prose, not lyric.

Now, anyone charged with evaluating something and reporting about it must also reveal at least some issues, lest it look suspicious. If handed the ten commandments under such a duty, we'd all find this or that commandment lacking in purpose or in need of clarification; in our best temper, we might limit our recommendation to reshuffling them. Something like this is to be expected in model validation as well—it will require some effort under the best of circumstances. The additional overhead will then vary greatly.

17.6   What About the Kurtosis?

It is likely that asset returns are not perfectly normally distributed, and it is generally assumed that returns exhibit tails heavier than normality would imply. A common way to measure this is the *kurtosis*. A normal distribution (of any standard deviation) has a kurtosis of 3. Larger kurtosis values denote heavier tails.

It is fairly easy to impose some kurtosis in a VaR model; one possible approach, that of scaling some of the joint returns, is given in Chap. 9. However, this would

first require having an idea about a desired or target kurtosis, and it is here that it gets a bit murky—for kurtosis is quite fickle a measure, and it may, worse, also overzealously indicate tails where there really are none.

Let's look at the volatility of this measure first, which can easily be observed by doing a small experiment in Excel. Just create 40 random normals and determine their kurtosis with Excel's KURT function repeatedly; the sample's kurtosis will fluctuate between roughly 2 and 4, even though the sample stems from a normal distribution with a kurtosis of 3. This variability is caused by the influential 4th power term used in computing the kurtosis. For the same reason, outliers in a sample may potentially massively increase the kurtosis. It is indeed not unusual to observe market data time series with a kurtosis of over 100. Now, comparing such potentially large and highly fluctuating kurtosis values is possibly deceptive—two kurtoses of 60 and 80 seem to differ a lot, for example, but can be caused by data sets that are quite similar in nature.

Can we obviate this by relying on larger samples? Unfortunately, in our setup of historical returns observed in markets whose volatility regimes seem to change over time, this can lead to an undesired signaling of tails—the kurtosis cries wolf. Why so? Consider a sample of 500 artificial normal returns (with a standard deviation of 1), whose sample kurtosis will hover around 3. Now replace the last 25% of those returns with new normal returns with a standard deviation of 2. Those new returns are also normally distributed, and their kurtosis will fluctuate around 3 as well. The full 500 artificial returns should resemble an asset whose price fluctuations have doubled in the last 6 months, as illustrated by the raw returns and their local volatilities in Fig. 17.2.

If we blindly compute the overall kurtosis, however, we obtain a value different from 3—say, 4.83. This is purely caused by mixing two different, innocent normals. But blessed with the knowledge of how this time series came about, we can dismiss this warning of non-normal, heavy tails: this is a time series of normals whose vola level changed but whose tail characteristics effectively remain normal. Such a time series simply would not call for worrying about or injecting kurtosis into our model.
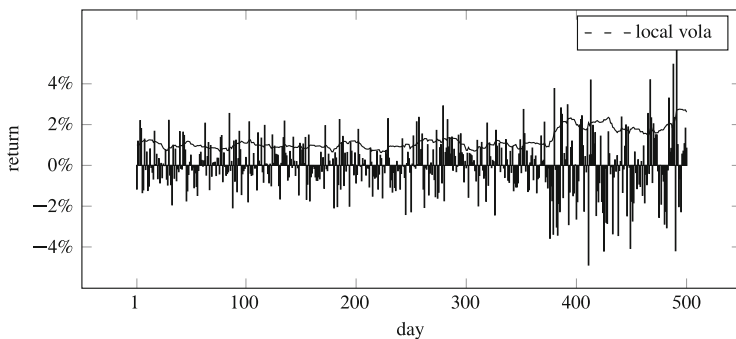

Fig. 17.2 Artificial returns with recently increased volatility

We can quantify this from a different angle: Fig. 17.3 depicts *local* kurtosis values over 40-day windows for the same time series. By their nature, the local kurtosis values fluctuate, but their median of 2.96 indicates no particular reason to believe in the large overall kurtosis.

We have seen that combining two sets of returns from different random normals will exhibit kurtosis. (Such combinations arise from *mixed* distributions, where a random variable can follow, at each realization, one of several distinct distributions). This is analytically tractable, but it is both simple and instructive to construct or simulate this effect. The border cases—taking all returns from the first normal distribution or taking all returns from the second one with a larger standard deviation—both result in a kurtosis of 3. Figure 17.4 shows how throwing an increasing percentage of large returns into a mix of normals affects the combined kurtosis. The volatile dotted lines are derived from 500 returns, where standard normal returns are replaced with more and more returns of standard deviation 2 (or 4). The smoother lines depict the same experiment done on a set of 100,000 normals.
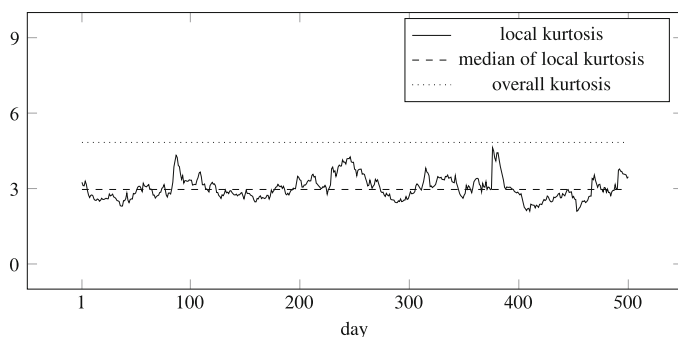

Fig. 17.3   Local kurtosis over time vs. overall one
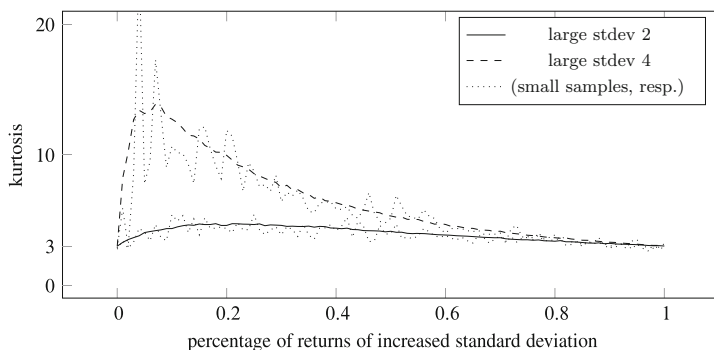

Fig. 17.4   Kurtosis of mixed normals

We first see that samples of size 500 are quite unstable with regard to kurtosis. More importantly, the kurtosis we experience—as exemplified in our artificial simulation—must be interpreted. If it is merely the artifact of vola regime changes, its tail indication might well be dismissed.

How would "real" kurtosis manifest itself? Let's revisit the same returns as in Figs. 17.2 and 17.3, but change their arrangement. Instead of putting all large returns at the end of the series (representing a vola regime change), we could arbitrarily mix them into the overall series, e.g., we could take every 4th overall return from the set of larger ones. First, obviously, the overall kurtosis isn't affected by the reshuffling and remains at 4.83. Yet this time series now is different—we changed its meaning. Now it represents (by our design) no longer a time series with recently increased vola; now it resembles one that frequently, intermittently exhibits shocks that are larger than expected. We can regard this as an actually heavy-tailed distribution as opposed to a mere sequence of two different normal ones. Under this interpretation of our series, and based on its new and known behavior over time, the overall kurtosis becomes meaningful (even though it has the same value).

Figure 17.5 shows what happens to our time series in this reshuffling exercise. Its overall kurtosis, as mentioned, is unchanged. But the behavior of the local kurtosis is very different now: all 40-day windows, and no longer only the right-most ones, now contain some large returns. Note how the median local kurtosis is no longer close to 3, as in the previous Fig. 17.3, but larger at 3.76.

Unfortunately and obviously, the picture is less clear-cut with real-world data. Figures 17.6 and 17.7 show returns and local kurtosis behavior for the Norwegian krone versus euro exchange rate. The local kurtosis is above the normal one in the first year; afterwards, it hovers closer to 3, indicating normality—except for two periods where return spikes (outliers?) cause brief periods of large local kurtosis. Even if we don't dismiss those two return instances as data flukes or political events, the overall kurtosis seems to be too large. Nevertheless, without the luxury of replications available in the artificial setup, where we can dismiss or at least corral in the kurtosis, agreeing on a target kurtosis for our one-off sample feels like an exercise with dubious success prospects.
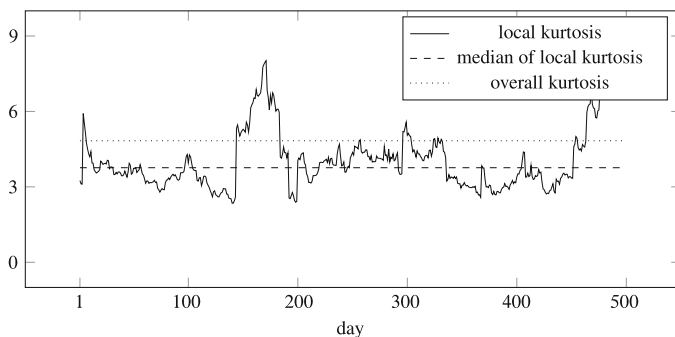

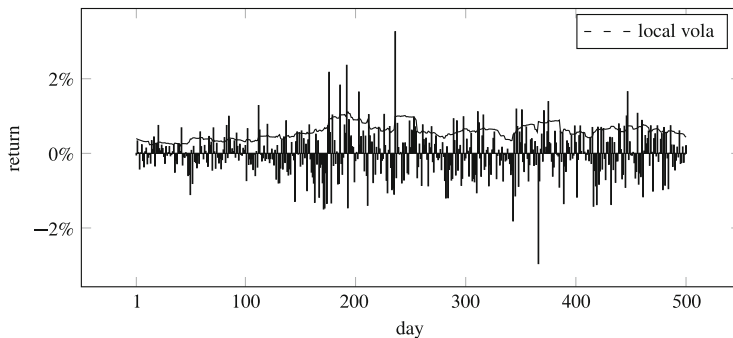**Fig. 17.5** Local kurtosis over time vs. overall one, with shuffled returns
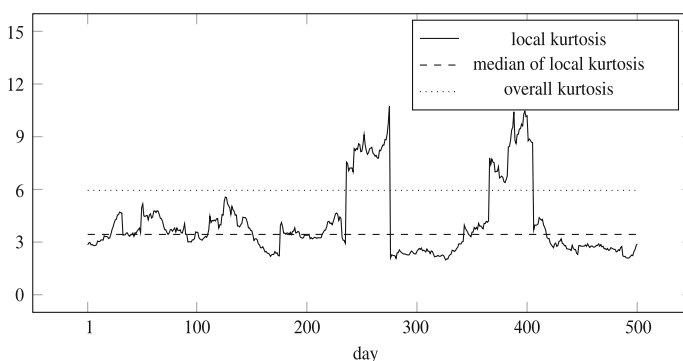
Fig. 17.6 Real returns

Fig. 17.7 Real local kurtosis over time vs. overall one

Distribution tails are certainly of interest, and an established measure for it is readily at hand with the kurtosis. The very interventionist operation of volatility rescaling may well reduce the overall historical kurtosis, as seen in our first example, and thus ostensibly suppress tails; a Monte Carlo approach applied on top of it would, by imposing normality, annihilate the kurtosis outright. This makes a discussion about it in the context of a VaR model almost inevitable.

What are the approaches that can and should be taken with regard to kurtosis, and how to defend them? By far the worst choice, in my opinion, is to tackle this issue in a conventional manner and try to model the kurtosis in detail, e.g., for each risk factor. Trying to pin down an actual kurtosis is a highly speculative exercise, due to spurious effects arising from vola level changes. In essence, we do not know the kurtosis. But even if one decided to ignore that and just blindly and technically computed the measure and imposed corresponding tails, the practical obstacles would be considerable: separate analyses on 2200 time series would be required; they would have to be repeated periodically in a dynamic VaR setup; and any kurtosis-related tweaks or optimizations would be unsteady as functions of a relatively fickle measure.

Possibly less elegant from a position of mathematical purity but far more practical is it to test the model via the standard backtesting and distribution tests. If the results are unconvincing and can be traced back to the modeling of tails, apply some simple heuristic to compensate for the shortfall; you could, for example, scale up the VaR by some constant factor or inject some fixed, artificial kurtosis as described in Chap. 9. This deliberately forgoes explicit criteria of optimality or kurtosis fits. It avoids lengthy validation discussions, prolonged by the underlying ignorance of any "real" kurtosis. It refuses to pretend to know the unknowable and avoids giving a false sense of confidence. It essentially puts the kurtosis topic outside the model's core realm.

On the other hand, if backtesting and distribution tests are able to plausibilize a model without explicit kurtosis handling, than ignoring kurtosis aspects is by far the preferable choice of action. In fact, one could view such end-to-end tests as strong indications that much of the technically signaled kurtosis is an artifact. We expect many real-world portfolios to perform just fine in a kurtosis-oblivious VaR model, and not only because often the main risk drivers happen to be liquid and quasi-normal assets like domestic interest rates and major foreign currencies. If the tests succeed, the benefits of further kurtosis tweaks surely appear to be marginal, especially in light of disproportionate efforts. In other words: if the distribution tests fit, you must acquit.

To sum up, we propose to neglect kurtosis issues whenever possible and justified by standard test procedures and to defend this decision on the case of equivocal kurtosis values in case of mixed distributions. If not viable, apply some simple catch-all heuristic to ballast the tails. Squash the topic or keep it at arm's length.