Equilibrium VIX in Inelastic Markets*

Albert J. Menkveld

October 21, 2025

Abstract

On average, the squared VIX exceeds realized variance. This implies that investors pay a premium to hold variance risk. But, why *pay* for risk? And, why does the premium correlate with volume? In Grossman-Miller type inelastic markets, investors hold variance risk to hedge against liquidity shocks, because these shocks cause price pressures that add to realized variance. Therefore, a positive variance risk premium must emerge in equilibrium. This result is developed formally, and the model is calibrated to match empirical patterns in the variance risk premium and trading volume around eleven crises between 1993 and 2025.

*Albert J. Menkveld, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands, albertjmenkveld@gmail.com. I thank Dion Bongaerts, Julio Crego, Bernard Dumas, Vincent Fardeau, Sergei Glebkin, and Jun Wu, and (seminar) participants at AFML 2024 at Corvinus U Budapest, CBCMFM 2024 at the Banco de Mexico, EDHEC, Helsinki GSB, and at Luiss Rome. I am for financial support from the Dutch Research Council (NWO) for grant number 016.Vici.185.068.

Equilibrium VIX in Inelastic Markets

Abstract

On average, the squared VIX exceeds realized variance. This implies that investors pay a premium to hold variance risk. But, why *pay* for risk? And, why does the premium correlate with volume? In Grossman-Miller type inelastic markets, investors hold variance risk to hedge against liquidity shocks, because these shocks cause price pressures that add to realized variance. Therefore, a positive variance risk premium must emerge in equilibrium. This result is developed formally, and the model is calibrated to match empirical patterns in the variance risk premium and trading volume around eleven crises between 1993 and 2025.

Empirically, being short variance has been shown to earn a positive premium, referred to as the variance risk premium (VRP). The premium can be rationalized by "asymmetric GARCH": Large *negative* returns make variance jump more than large positive returns do. Investors can, therefore, hedge against large negative returns by being long variance in a variance swap. Suppose they enter the swap at the start of the month to receive realized variance (RV) at the end of it, where RV is defined as the sum of squared daily returns. Asymmetric GARCH implies that high RV states tend to be states of large negative returns, hence the hedge. The price investors pay to be long variance swaps is referred to as implied variance (IV). The hedging motive implies that the swap market can only clear if IV exceeds expected RV, hence a positive VRP.[1] The squared VIX was constructed in such a way to approximate IV, thus linking VRP to the level of VIX.[2]

The exogenous asymmetry that these explanations rely on is critical in the sense that, without it, the variance risk premium cannot be signed from first principles. The VRP does not follow from Jensen's inequality, which is used to sign the equity risk premium. The logic for the latter is that the expected utility of (stochastic) dividend is less than the utility of expected dividend, because utility is a concave function. The equity risk premium, therefore, is nonnegative. This logic no longer applies when the expectation is taken over a composite function: concave utility and convex squaring. The expected utility of squared dividend can, therefore, not be signed relative to the utility of squared expected dividend.

I propose a model that yields a positive variance risk premium without relying on exogenous asymmetry. It is generated endogenously in a symmetric economy with a liquidity risk friction. I use the baseline model in Vayanos and Wang (2012) (VW12), which is a Grossman-Miller

[1] A well cited literature explains the positive VRP this way (see, e.g., Bakshi and Kapadia (2003), Bollerslev, Tauchen and Zhou (2009), or Todorov (2010)). Bekaert, Engstrom and Ermolov (2023) explains VRP patterns by an asymmetry in consumption growth uncertainty.

[2] Carr and Wu (2006) and Carr and Wu (2009) discuss how the construction of VIX by the Chicago Board Options Exchange (CBOE) was done in such a way that the squared VIX is a model-free approximation of implied variance. Appendix A summarizes their derivation.

type of model. The model features three periods. Risk-averse investors are identical in the initial period. In the second period, a fraction of them experiences an identical[3] liquidity shock[4]. These investors seek to hedge the shock by trading the asset, and thus become liquidity demanders. The nonshocked investors become the *de facto* liquidity suppliers. The price becomes (temporarily) pressured to pay these suppliers a liquidity premium. Hence, the market is inelastic. In the final period, investors consume.

The only change to this baseline VW12 model is that a variance swap is added that trades in first period, and pays off in the third period. The short side of the swap pays (stochastic) RV to the long side in the final period. In return, they receive a pre-agreed (deterministic) payment at maturity: IV. As the swap is in zero net supply, the market only clears in the first period if the implied variance equals the certainty equivalent of future RV. In the CARA-Normal VW12, this expectation is analytic and thus allows for studying the drivers of the variance risk premium.

The model yields the following insights. First, VRP is positive in the presence of liquidity risk. The intuition for this result is that the expected utility of receiving RV for an investor in the first period is a weighted average. With probability κ, the investor becomes liquidity demander, in which case the expected utility of receiving RV is particularly high for two reasons. First, the shock reduces his wealth and, therefore, marginal utility increases. Second, large shocks imply having to pay a large (transitory) price pressure to compensate liquidity suppliers. This pressure adds to RV and thus makes *receiving* RV a hedge against large shocks. With probability $(1 - \kappa)$, the investor becomes liquidity supplier and thus experiences the reverse effects. However, it turns out that the effects are stronger for liquidity demanders, thus yielding a positive VRP. The effects being stronger for shocked agents as opposed to

[3]I conjecture that the insights that the model offers do not hinge on the shocks being identical. Positively correlated shocks across the fraction of shocked investors create the nonzero price pressure that drives the results.

[4]VW12 refers to this shock as a *nontraded risk shock*, which is more precise. I prefer to use liquidity shock to emphasize that it is liquidity risk that the variance swap is a hedge for.

nonshocked agents is the result of utility being concave: Marginal utility increases more for negative shocks than it decreases for positive shocks.

Second, VRP increases in the size of the friction, both on the intensive and on the extensive margin. That is, VRP increases if shocked investors experience larger shocks, or if more investors experience the shock, respectively. These two effects can be identified empirically by tracking both illiquidity and volume. A calibration to real-world data provides an illustration.

Third, if the liquidity shock correlates with the change in the state of the economy, then this amplifies VRP. An example of such correlation is that leveraged investors might be forced to sell on poor performance, or could lever up and buy on good performance (e.g., Vayanos, 2004). This mechanism is straightforward to implement in the model by revealing part of the dividend in the second period. The leverage-induced trading can then be modeled by making the liquidity shock correlate negatively with the dividend news. The hedging trade then makes investors buy on good news and sell on bad news. This amplifies VRP, essentially because the wealth effect becomes more severe.

Calibration. The model is illustrated by calibrating it to 11 crises from the start of VIX in 1993 through 2025. It has been documented that the VRP becomes elevated in the months after crises (Bekaert and Hoerova, 2014; Choi, Mueller and Vedolin, 2017). Both IV and RV jump during at the onset of a crisis, but IV decays more slowly than RV. Two explanations for this rather anomalous pattern have been offered recently. The first one is a behavioral one centered on investors who are slow to learn the decline in RV after a crisis (Lochstoer and Muir, 2022). A second one relies on heterogeneity in beliefs where agents disagree more about the state of the economy after a crisis (Martin and Papadimitriou, 2022).

The proposed model can rationalize the increased wedge between IV and RV post-crisis by stronger liquidity shocks. Comparing the calibrated values for the pre- and post-crisis period reveals that, on the extensive margin, fewer investors need to trade, but, on the intensive margin, they trade a lot more. This matches not only pre- and post-VRP levels, but also the unusual pattern of 81% more volume (in shares) *and* 26% higher illiquidity (realized bid-ask spread) in the post-crisis period.[5] Investors who seem to trade more in illiquid markets post-crisis might be under pressure to delever, and thus, in a sense, have no choice but to trade.[6] As this pressure to delever might hit anyone at the start of these post-crisis months, IV and, therefore, VIX remains elevated in the post-crisis months in spite of quickly dropping RV. The calibration shows that this leverage-induced trading channel likely is nonnegligible, because about a half of the variance of long-term changes to the S&P500 index can be explained by daily net volume in SPY, which is an actively traded exchange-traded fund (ETF) that tracks it.

Related literature. Recent studies on the pricing of VIX have focused on market makers. Cheng (2019), for example, shows that the variance exposure of CBOE market makers in VIX futures and index options is, on average, negative. Apparently, they are unable to fully hedge their net variance exposure. Fournier and Jacobs (2020) offer a model that generates a positive VRP featuring a market maker with limited capital. The proposed model adds to this literature by endogenizing liquidity supply and thus adds a general-equilibrium perspective.

Another related study on liquidity and volatility is Drechsler, Moreira and Savov (2021). They identify another channel by which liquidity suppliers are negatively exposed to variance shocks. In a nutshell, an unexpected jump in variance hurts them, because it comes with more information asymmetry than they priced in when submitting their supply schedules.

[5] Konstantinidi and Skiadopoulos (2016) find that among four candidate models to explain VRP, the one with trading activity best predicts it out of sample. This suggests that there is an important role for trading in a model that prices variance swaps.

[6] The long-term price response to leverage-induced trading could be endogenous in the following sense. Markets are inelastic, both in terms of commanding a short- *and* a long-term price response to net volume. Evans and Lyons (2002) document this finding for foreign exchange trading and interpret it as due to "portfolio shifts" by some investors that need to be absorbed by other risk-averse investors. Recently, Gabaix and Koijen (2024) propose the "inelastic market hypothesis" that relates inelasticity to limited flexibility of institutions when responding to asset price fluctuations. Gârleanu, Pedersen and Poteshman (2009) discuss inelastic markets in the context of derivatives.

They, therefore, suffer more adverse-selection costs than provided for in their quotes, which leads to unanticipated losses. Vice versa, unexpected variance declines lead to surprise profits. This channel is complementary to the liquidity-risk channel that links illiquidity directly to the VIX level.

The rest of the paper is organized as follows. Section 1 presents the model and derives various theoretical results. Section 2 calibrates the model to match VRP and trading patterns around crises. Section 3 concludes.

1 Model

The variance swap is priced in a Grossman-Miller model with liquidity shocks. More specifically, the proposed model is the baseline model in Vayanos and Wang (2012), henceforth referred to as VW12. The root cause for orders suffering (transitory) price impact is risk aversion. The model is, therefore, in the spirit of Grossman and Miller (1988).[7] In a nutshell, the proposed most general model extends VW12 in two ways. First, the financial market is extended with a derivative, a variance swap, which enables me to study equilibrium pricing of this derivative in a relatively standard setting. Second, all agents receive a dividend signal in the intermediate round, which creates a need to unwind or enlarge positions.

1.1 Model primitives

[Figure 1 about here.]

Figure 1 summarizes the model. It features three periods, indexed by $t \in \{1, 2, 3\}$. The financial market consists of a risky asset and a variance swap. The risky asset is in supply of

[7]The other root cause for price impact is information asymmetry as in Glosten and Milgrom (1985). Information is symmmetric in the baseline VW12 model.

$\bar{\theta}$ shares that pay v units of a consumption good in Period 3, where v is Gaussian with mean μ_v and variance σ_v^2. The price of this asset at time t is p_t.[8]

The variance swap is in zero net supply. It trades in Period 1 and pays off the realized variance of the risky asset in the final period:[9]

$$V = \sum_{\tau=2}^{3} (p_\tau - p_{\tau-1})^2 . \tag{1}$$

The expected realized variance is denoted as:

$$x_1 = E(V) . \tag{2}$$

The price of this swap, y_1, is such that the Period 1 investor is indifferent between receiving a (deterministic) payment y_1, or receiving the (stochastic) realized variance at Period 3:

$$y_1 = E(V \times m_3) , \tag{3}$$

where m_3 is the (state-dependent) marginal utility of an agent in the final period.[10] This equality that pins down the price of swap is a direct result of the derivative being in zero net supply and agents being homogeneous *ex-ante*. Finally, note that the swap does not trade in Period 2. This is in line with the asset pricing literature that is focused on the payoff to variance swaps entered into at the beginning of the month, and paid out at the end of it (e.g., Carr and Wu, 2009).

[8] All prices are expressed in units of Period 3 consumption good. In a more extensive model, this is the outcome of introducing a riskless asset that is used as numéraire for all asset prices (e.g., VW12). I prefer to keep the presentation brief.

[9] Price differentials in this model are referred to as returns. All empirical analysis in this manuscript uses log prices, so that differentials can indeed be interpreted as returns.

[10] Note that the expression for y_1 does not feature $E_{t=1}(m_3)$ in the denominator. The reason is that prices are expressed in terms of the Period 3 consumption good as discussed in footnote 8.

Agents. The model features a measure one of agents. Agents are risk-averse and derive utility from consumption in Period 3. Their utility over the consumption good is exponential:

$$U(c_3) = -\exp(-\alpha c_3), \tag{4}$$

where $\alpha > 0$ is the coefficient of absolute risk aversion. Agents are identical in Period 1 and are, therefore, endowed with the per-capita supply of the asset and the derivative.

Trade is generated in Period 2 as follows. In this period, agents observe part of the dividend. They will learn the remainder in Period 3. Total dividend can therefore be written as:[11]

$$v = v_2 + v_3, \tag{5}$$

where

$$v_2 := E_{t=2}(v), \quad v_2 \sim N\left(\mu_v, \sigma_{v_2}^2\right), \tag{6}$$

and

$$v_3 \sim N\left(0, \sigma_{v_3}^2\right), \quad \sigma_{v_3}^2 = \sigma^2 - \sigma_{v_2}^2, \tag{7}$$

where

$$\sigma_v^2 = \sigma_{v_2}^2 + \sigma_{v_3}^2. \tag{8}$$

After all agents observe v_2, a fraction κ of them receives the liquidity shock z, which implies a payoff of $(z \times v_3)$ of the consumption good in Period 3. These agents observe z with:

$$z \sim N(0, \sigma_z^2), \quad \text{Corr}(v_2, z) = -1, \quad \text{Corr}(z, v_3) = 0. \tag{9}$$

The agents who experience the shock $(z \times v_3)$ can hedge it by trading the risky asset. For $z > 0$, for example, these agents can hedge this risk by selling the asset. These agents,

[11] One way to generate this structure is to let agents receive a noisy signal on the dividend in Period 2 as in Appendix C.

therefore, initiate trades in Period 2 and become liquidity demanders. The remaining agents accommodate these trades and thus become the *de facto* liquidity suppliers. Although the model is based on two types only, it is worth quoting the opening line of Dumas (1998):

> The two-investor equilibrium is as basic to financial economics as is the two-body problem to mechanics.

The perfectly negative correlation between v_2 and z could be interpreted as leverage-induced trading in the following sense. Bad news on the dividend coincides with positive z and, therefore, makes shocked investors sell. Conversely, good news makes them buy. This is consistent with investors holding leveraged positions, which they might be forced to liquidate if the value falls below a threshold (e.g., Vayanos, 2004).

CARA-Normal models have become workhorse models in the literature, because they often yields tractable results. The same is true for the proposed model. With normality, however, the shock $(z \times v_3)$ can take extremely large negative values, which can make expected utility infinitely negative. To avoid this corner, VW12 impose the following condition on the variance of v_3 and z (c.f., VW12, Eqn. (1.2)):

$$\alpha^2 \sigma_v^2 \sigma_z^2 < 1. \tag{10}$$

1.2 Equilibrium quantities

This section presents equilibrium expressions for the quantities in the flow chart at the bottom of Figure 1. They are derived working backwards from the final period of the model. All proofs are in Appendix D and key derivations have been double-checked by means of a Mathematica notebook that is available at `https://bit.ly/3CM8lDZ`.

Assumption 1 ***(Zero intercept.)*** *To focus the analysis on the risk transfer only and thus avoid nonzero intercept terms in returns, it is assumed that $\bar{\theta} = \mu_v = 0$ in the remainder.*

Assumption 1 allows us to ignore intercept terms when pricing variance (as discussed below, following Lemma 2). This keeps expressions clean and, thereby, maximizes economic insight.[12] With Assumption 1, the security becomes a vehicle for investors to transfer *only risk.* For completeness, the proofs in Appendix D feature closed-form expressions for the general case of nonzero $\bar{\theta}$ and μ_v.

1.2.1 Period 3

The final period quantities are trivial. The price of the risky asset equals the dividend realization: $p_3 = v_2 + v_3$. The payoff to being long the variance swap equals the squared realized price differentials as per (1).

1.2.2 Period 2

The most important feature of Period 2 is trading in the risky asset. This trading yields expressions for the following three quantities: A market-clearing price p_2, net volume q_2 which is equal to $\Delta\theta_2^d$, and price impact λ_2. Since liquidity demanders initiate trades, their demand informs the sign of net volume: negative if they are selling, positive if they are buying. A standard illiquidity measure is "Kyle's λ." It is the regression coefficient that results from regressing price change on net volume (i.e., regressing $(p_2 - p_1)$ on q_2).

Lemma 1 ***(Period 2 equilibrium quantities.)*** *This lemma presents equilibrium quantities for trading in the risky asset in Period 2:*

$$p_2 = v_2 - \alpha\sigma_3^2\kappa z, \qquad \textit{(Price)} \tag{11}$$

$$q_2 = \Delta\theta_2^d = -\kappa\,(1-\kappa)\,z, \qquad \textit{(Net volume)} \tag{12}$$

[12] Assumption 1 is relatively harmless for the analysis of equilibrium VIX, which is based on the sum of squared *daily* S&P500 index returns. It, therefore, focuses on the second moment, which is equal to the sum of the squared mean and the variance. As the squared mean is an order of magnitude smaller than variance for daily market returns, the zero-intercept assumption is, indeed, largely inconsequential.

$$\lambda_2 = \underbrace{\frac{\sigma_{v_2}}{\kappa\,(1-\kappa)\,\sigma_z}}_{\substack{\textit{Permanent}\\ \textit{price impact}}} + \underbrace{\frac{\alpha\sigma_{v_3}^2}{1-\kappa}}_{\substack{\textit{Transitory}\\ \textit{price impact/}\\ \textit{price pressure}}}\ . \qquad \textit{(Illiquidity)} \qquad (13)$$

The results of Lemma 1 can be understood as follows. The price equation shows that p_2 is equal to the expected dividend v_2, plus (transitory) price pressure caused by the liquidity shock z. For example, for a positive value of the shock, liquidity demanders sell, and p_2 settles below v_2 to compensate liquidity suppliers for taking the other side.[13]

The net volume equation scales negatively with z, because hedging z requires an offsetting position in the asset. The size of the scaling factor reaches its maximum at $\kappa = 1/2$. The intuition is that values other than a half create an imbalance in the presence of demanders and suppliers. The further κ is below a half, the fewer demanders there are relative to suppliers, which constrains trade opportunities. The same holds for values above a half, but now driven by fewer suppliers relative to demanders. This all leads to the interesting result that an increase in the fraction of agents who seek liquidity leads to more volume, but only up to a point, after which volume declines. Volume as a function of κ is hump shaped.

Finally, the illiquidity equation shows that price impact consists of two canonical components: A permanent and a transitory one. The permanent component is the immediate result of z correlating negatively with v_2, thus making net volume correlate positively with v_2. This not a causal relationship, but a reflection that both are jointly determined in equilibrium, potentially reflecting leverage-induced trading.[14] The transitory component is entirely due to net volume correlating positively with the price pressure term in p_2. This correlation is positive, again because net volume correlates negatively with z.

[13]Note that there is no additional term to reflect an equity risk premium in (13). The reason is that Assumption (1) states that the total supply of the asset, $\bar{\theta}$, is zero. An equity risk premium is there for the more general case of $\bar{\theta} \geq 0$, which is in (42) of Appendix D.

[14]This result squares well with the empirical pattern of net volume correlating substantially with long-term index changes (see, e.g., Gabaix and Koijen, 2024). This pattern also shows up in the calibration of Section 2.

1.2.3 Period 1

Before being able to price the variance swap, the Period 1 price of the risky asset is needed.

Lemma 2 ***(Period 1 price risky asset.)*** *The price of the risky asset in Period 1 is:*

$$p_1 = \mu_v = 0. \tag{14}$$

Lemma 2 states that the price of the asset in Period 1 is equal to the dividend mean: μ_v. Again, there is no equity risk premium for the same reason that there is no such premium in p_2 (see earlier discussion).

Main theoretical results. With these intermediate results in place, it is time to state the main theoretical results. They center on the price of the variance swap in Period 1. Recall that the expression for realized variance is:

$$V = (p_2 - p_1)^2 + (p_3 - p_2)^2 . \tag{15}$$

The long side of the swap gets paid this variance in Period 3. In return, this side needs to pay the short side of the swap a fixed amount at maturity, which is referred to as the price of the swap: y_1. Since the swap is in zero net supply, its price in Period 1 must be equal to the reservation value of the agent. This value is well defined, because all agents are identical in Period 1.

Proposition 1 ***(Period 1 price variance swap.)*** *The Period 1 price of the variance swap is:*

$$y_1 = \kappa y_1^d + (1 - \kappa)\, y_1^s =$$

$$= \underbrace{\left(\kappa A^2 + (1-\kappa) B^2\right)}_{\substack{\textit{Inflator flow-correlated} \\ \textit{return variance}}} \times \underbrace{\left(\sigma_{v_2} + \alpha \sigma_{v_3}^2 \sigma_z \kappa\right)^2}_{\substack{\textit{Flow-correlated} \\ \textit{return variance}}} +$$

$$+ \underbrace{(\kappa A + (1-\kappa) B)}_{\substack{\textit{Inflator flow-uncorrelated} \\ \textit{return variance}}} \times \underbrace{\sigma_{v_3}^2}_{\substack{\textit{Flow-uncorrelated} \\ \textit{return variance}}}, \tag{16}$$

with

$$A = \frac{1}{1 - \alpha^2 \sigma_{v_3}^2 \sigma_z^2 (2-\kappa) \kappa} \geq 1 \quad \textit{and} \quad B = \frac{1}{1 + \alpha^2 \sigma_{v_3}^2 \sigma_z^2 \kappa^2} \leq 1. \tag{17}$$

The following inequality characterizes the two inflators:

$$\kappa A^2 + (1-\kappa) B^2 \geq \kappa A + (1-\kappa) B \geq 1. \tag{18}$$

Proposition 1 presents the closed-form expression for the price of the variance swap in Period 1. The first line of (16) merely states that, for an agent in Period 1, it is equal to his expected value whereby, with probability κ, he receives a shock, in which case his expected value of being long the swap is y_1^d, and, with probability $(1-\kappa)$, he does not receive any shock and, endogenously, becomes liquidity supplier whose expected value of being long the swap is y_1^s.

[Figure 2 about here.]

The next two lines in (16) express the value of the swap in terms of the model parameters. The expression is intuitive in the sense that implied variance is equal to expected realized variance, whereby the two components of the latter are inflated by well-defined factors, referred to as inflators. The two components are one that is correlated with net volume, i.e., with z, and one that is orthogonal to it.[15] The inflators for these two components look very similar. The only difference is that the inflator for the flow-correlated component involves squared factors. The proposition states that both inflators are weakly larger than

[15]Note that, for the flow-correlated part, (transitory) price pressure is added to the permanent price impact. This captures prices overshooting fundamental value changes to pay for liquidity supply. Be reminded that net volume $q_2 = (-\kappa(1-\kappa)) z$ correlates positively with v_2, because z correlates negatively with it.

one, and, therefore, are true *inflators*. It further states that the flow-correlated inflator weakly dominates the flow-uncorrelated one. Figure 2 illustrates this finding.

The results in Proposition 1 imply that the variance risk premium cannot be negative. The model therefore delivers nonnegativity of the VRP, which is a nontrivial result as it does not follow from first principles. As argued in the introduction, Jensen's inequality cannot be invoked to sign the premium that risk-averse investors require for holding *squared* dividend. The nonnegativity result that immediately follows from Proposition 1 is, therefore, an important result in explaining why empirical studies consistently find a positive variance risk premium. To emphasize its importance, the result is stated as a corollary:

Corollary 1 ***(Variance risk premium nonnegative.)*** *The variance risk premium is nonnegative:*

$$y_1 - x_1 \geq 0. \tag{19}$$

To develop more insight into what drives the price of the variance swap, it is worth exploring whether partial derivatives can be signed. The next proposition develops three useful results in this respect.

Proposition 2 ***(Variance risk premium monotonicity.)*** *The variance risk premium,* $y_1 - x_1$*, increases monotonically in*

- *the fraction of agents who receive the liquidity shock* (κ)*,*
- *the level of risk aversion* (α)*, and*
- *the size of the liquidity shock* (σ_z^2)*.*

Proposition 2 finds that VRP increases monotonically in κ, in α, and in σ_z^2. The latter two are intuitive in the sense that more risk to be hedged, or a higher risk aversion, make the variance swap more valuable in terms of hedging against shock uncertainty, and the VRP therefore rises.

[Figure 3 about here.]

The result that the VRP increases in κ is, *a priori*, less intuitive. Figure 3 illustrates this finding by plotting x_1, y_1^d, and y_1^s against κ. The figure provides economic intuition for the key insight of the paper: VRP being nonnegative in the presence of liquidity risk.

First, the expected realized variance, x_1, increases in the fraction of shocked agents (solid blue line). The reason is that more liquidity demand commands higher price pressure, which adds to overall expected realized variance:

$$x_1 = \underbrace{\left(\sigma_{v_2} + \alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2}_{\substack{\text{Flow-correlated} \\ \text{return variance}}} + \underbrace{\sigma_{v_3}^2}_{\substack{\text{Flow-uncorrelated} \\ \text{return variance}}} . \tag{20}$$

The term that is added to σ_2 before squaring is the standard deviation of (transitory) price pressure, which, as per (13) and (12), is:

$$\sigma_{q_2\times\lambda_2} = \alpha\sigma_{v_3}^2\kappa\sigma_z. \tag{21}$$

Second, for shocked investors, who become liquidity demanders, the expected *value* of receiving realized variance exceeds expected realized variance (dotted red vs. solid blue line). The market is incomplete in the sense that liquidity risk cannot be perfectly hedged *ex-ante*. The liquidity shock, therefore, adds risk which is costly to a risk-averse agent. Being long a variance swap undoes some of this utility cost, because the utility loss caused by large (in magnitude) shocks coincides with higher payoff on realized variance. In other words, the larger price pressures paid by shocked investors coincide with higher payoffs on the swap, simply because these larger price pressures add to realized variance. Note that this economic channel is driven by liquidity risk, not by leverage-induced trading. This observation immediately follows from setting σ_2^2 to zero in Proposition 1 and Corollary 1.

Third, for nonshocked investors, who become liquidity suppliers, the expected value of receiving realized variance is below expected realized variance (dot-dash green vs. solid blue line). The reason is that all effects discussed for shocked agents also apply to nonshocked agents, but, rather than negative wealth shocks, they serve as positive wealth shocks. Therefore, higher payoffs on realized variance coincide with *higher* as opposed to *lower* wealth states. These payoffs, therefore, amplify wealth shocks for nonshocked agents as opposed to dampen them. This makes nonshocked agents appreciate receiving realized variance less than its expected value.

Fourth, the value of the variance swap is the weighted average of its value to shocked and to nonshocked investors. This weighted average exceeds expected realized variance (orange dashed vs. solid blue line). The wedge between the two increases in the fraction of shocked agents, which illustrates the formal result stated in Corollary 1. The reason for this finding is that, when increasing the fraction, the expected *positive* hedging value of variance swaps in shocked states increases relative to the expected *negative* value in nonshocked states. This makes sense for concave utility investors, whose marginal utility increases more in negative states than it decreases in positive states.

Leverage-induced trading. The key insight of the paper, VRP being nonnegative in the presence of liquidity risk, does not rely on the correlation of net volume and permanent price change in the intermediate period, interpreted as leverage-induced trading. Therefore, the model could have been smaller by omitting it. There are two reasons for keeping this leverage-induced trading in the model. First, the calibration to real-world VIX and index trading benefits from adding leverage-induced trading to the model. A substantial part of monthly realized variance, about a half, is explained by net volume correlating with (long-term) index changes. Second, the result in Proposition 1 provides an interesting comparative statics result. If more of the fundamental value correlates with net volume, i.e., if $\sigma_{v_2}^2$ is a larger part of σ_v^2, then VRP is higher. Leverage-induced trading, therefore, is itself a driver of the VRP level and, therefore, VIX. It is worth stating this insight formally as a corollary:

Corollary 2 ***(Leverage-induced trading amplifies the variance risk premium.)*** *Consider the case of positive liquidity risk ($\kappa > 0, \sigma_z > 0$) and fix total fundamental variance at σ_v^2. If one increases the leverage-induced trading component of fundamental variance, then VRP increases. Formally,*

$$\frac{\partial}{\partial \sigma_{v_2}} (y_1 - x_1) > 0, \tag{22}$$

where

$$\sigma_{v_3}^2 = \sigma_v^2 - \sigma_{v_2}^2. \tag{23}$$

The proof immediately follows from (16) and (18).

2 Calibration

To illustrate the model, it is calibrated twice to match pre- and post-crisis trading. Can the model rationalize a jump in the VRP post-crisis? And, can it rationalize *more* trading on *less* liquidity post-crisis? And, more importantly, what are the economic channels that cause these patterns? This section first establishes the empirical patterns around crises, and then turns to calibrating of the model to the pre- and the post-crisis pattern.

2.1 Empirical patterns

This subsection presents several empirical results. It first confirms the variance patterns that have been documented in the literature, both for RV and IV. It does so for a longer sample that includes 11 crisis periods. It then complements these pricing patterns with trading patterns based on intraday trade data.

2.1.1 Sample

The sample is constructed in such a way that it enables calibration of the model to the VIX. It requires data on the VIX as a proxy for the squared IV, the S&P500 index (SPX), and trading in this index. The latter is proxied for by trade data for SPY, which is an ETF that tracks SPX. SPY is among the most actively traded ETFs and can, therefore, be considered broadly representative of index-based trading activity.

The sample starts on February 4, 1993, and ends on September 11, 2025. The start of the sample coincides with the start of SPY. 1993 was also the year when CBOE officially introduced the VIX. The sample, therefore, can not be extended further back in time.

The sample is sourced from Wharton Research Data Services (WRDS) TAQ, CBOE, and Nasdaq. The following time series are constructed based on these data sources: lp0.6 • *SPX* End-of-month (EOM) value of the S&P500 index.

• *Realized variance (RV)* EOM sum of squared daily SPX returns.

• *Implied variance (IV)* Beginning-of-month (BOM) squared VIX.

• *Realized VRP* BOM implied variance minus EOM realized variance.

• *Volume* EOM SPY volume in shares.

• *Effective spread* EOM share-weighted effective spread in basis points.

• *Realized spread* EOM share-weighted realized spread in basis points.

Details on the construction of all variables used in the analysis are in Appendix E.

[Table 1 about here.]

2.1.2 Exploratory analysis

This subsection presents various empirical results. First, it analyzes the IV, the RV, and the VRP time series to reproduce stylized facts. Second, it produces additional results based on trade data for SPY. This evidence is novel. Third, it zooms into crisis periods to characterize pre- to post-crisis patterns in IV, RV, VRP, and trading in SPY.

Time series for IV, RV, and VRP. Table 1 presents summary statistics for the three key variance-based variables: implied variance, realized variance, and the variance risk premium. These statistics serve as a "reality check," because they can be compared to the ones reported in Carr and Wu (2009, Table 2). My sample extends theirs in the sense that it covers 33 years (1993-2025) instead of the eight years they analyze (1996-2003). The summary statistics lead to the following observations. First, the average realized VRP is positive. That is, the average end-of-month realized variance exceeds the beginning-of-month implied variance (i.e., squared VIX). Therefore, those who shorted the variance swap at the beginning of the month earn a premium. This premium is sizable, because it amounts to $101/342 = 30\%$ annually.

[Figure 4 about here.]

Second, all three series are positively autocorrelated. The autocorrelation is particularly strong for implied variance: 0.76. The autocorrelation for realized variance is 0.48. The wedge between these two can be explained by the nature of the spikes in RV and IV. Figure 4 shows that the spikes are higher for RV, and that they drop more quickly. This quicker decay for RV is the first evidence of the pattern documented in the literature that the VRP remains elevated in post-crisis periods. The figure further identifies the 11 crisis months that correspond to the highest peaks, including, for example, the global financial crisis and Trump liberation day. I will zoom in on these crisis months when calibrating the model in Section 2.2.

These observations are very similar to those documented by Carr and Wu (2009). The premium for shorting variance is higher in their sample, but skewness, kurtosis, and autocorrelation are somewhat higher in my sample. Importantly, the time series features are the same in both samples.

[Figure 5 about here.]

Time series SPY trading. Figure 5 depicts how SPY trading evolves throughout the sample. It plots volume along with two standard liquidity measures: the effective spread and the realized spread. The effective spread measures the distance between the log transaction price and the log prevailing midquote, where midquote is defined as the average of the best bid and ask quote. This distance is a measure for the price impact that a liquidity demander effectively paid. The transitory component of this price impact is referred to as the realized spread.[16]

The figure shows a few interesting patterns. SPY volume grows steadily since the inception of SPY, but peaks at around 2015. It seems to stabilize at approximately one billion shares per month since then. It is not surprising that the two spread measures show the inverse pattern, as spread tends to be low when volume is high. The figure further suggests that there is a particular pattern for volume and liquidity around crises. Volume seems to jump in crisis months, to remain elevated in the months after. This volume increase is accompanied by *worse* liquidity in the sense of a higher spread, both effective and realized. This pattern is somewhat surprising given the general negative correlation between volume and spread. Whether these crisis patterns are real or merely noise, is analyzed next.

[Figure 6 about here.]

Crisis patterns. Figure 6 plots the various variance and trade quantities around the 11 identified crises. This done by bucketing the quantities by month relative to the crisis, and then computing the average for each bucket. The top three plots show how IV, RV, and VRP develop in crisis periods. RV jumps sharply in the crisis month, but declines quickly to the pre-crisis level in the three months after. IV, on the other hand, jumps in the month after the crisis, and decays relatively slowly to the pre-crisis level. Note that the one-month delayed

[16]The transitory component is identified by the extent to which the midquote falls short of completely moving towards the transaction price. The data in this paper are from *WRDS Intraday Indicators*, which computes these measures following the (standard) approach used in Holden and Jacobsen (2014) (which identifies the long-term midquote by the midquote prevailing five minutes after the trade.

jump is an artefact of IV being a *beginning-of-month* variable, as opposed to RV, which is an end-of-month variable.[17] The different levels of decay explain why the VRP jumps in the month after the crisis, and remains elevated in subsequent months. These patterns are statistically significant as judged by the confidence intervals.

The bottom three graphs in the figure illustrate the trading patterns. Volume more than doubles in the crisis month, and remains about one hundred percent higher subsequent months. The effective spread seems to double in the crisis month, to slowly decay in the post-crisis months. The realized spread shows the same pattern, although slightly less pronounced. All these trading patterns are not only substantial in economic terms, they are also statistically significant.

2.2 Calibration to crisis pattern

The model is calibrated to match the crisis patterns in VRP and trading. The calibration focuses on four parameters to match these patterns: Pre- and post-crisis values for the fraction of agents who experience the shock (κ), and the size of the shock (σ_z). These parameters are picked to match the following four moments in the data:

1. Pre-crisis VRP,
2. Post-crisis VRP,
3. Relative change in the (transitory) price pressure from pre- to post-crisis, and
4. Relative change in volume from pre- to post-crisis.

Note that a perfect match is not guaranteed, because the four empirical moments might lie outside of the space generated by the model-implied moments.

[17]The rationale for this timing throughout the paper is that one-month variance swaps are entered into at the start of each month, and paid out by the end of it.

The calibration further needs the following information: Pre- and post-crisis realized variance, decomposed in a flow-correlated and a flow-uncorrelated part, and the level of risk aversion. The latter is often assumed to be in the range from one to five, so it seems natural to set α to three. This level implies a reasonable risk premium for holding the market portfolio.[18] The calibration further relies critically on the monotonicity results of Proposition 2. Monotonicity ensures that, if there is a match, then there is a unique set of parameters that delivers this match. This desired monotonicity is also the reason for matching (transitory) price pressure, instead of total price impact. (13) shows that this price pressure is monotonous in κ, while total price impact is not. The details of the calibration procedure are in Appendix F.

[Table 2 about here.]

Table 2 presents the results of the calibration. Panel (a) quantifies the pattern to be matched, as depicted in Figure 6. The VRP jumps by a factor of 17 from the pre- to the post-crisis period, from $373 - 354 = 19$ percent squared to $1189 - 870 = 319$ percent squared. The realized spread, which proxies for the (transitory) price pressure, increases by 26%, and volume increases by 81%.

The calibration further takes as input the level of the two components of realized variance: the flow-correlated and the flow-uncorrelated component. This additional information is in panel (c). Realized variance increases from 354 to 870 percent squared, an increase of 146%. The VRP increases, because implied variance increases by a larger factor: $1189/373-1 = 219\%$. The flow-correlated component accounts for 51% of pre-crisis realized variance. It drops slightly to 44% in the post-crisis period.[19] Finally, panel (a) shows that the model is indeed able to match the empirical patterns perfectly.

[18]A CARA investor requires a risk premium of $0.5 \times 3 \times 0.20^2 = 6\%$ for holding a market portfolio with a volatility of 20%. These values are in line with the risk and return on the US market portfolio in recent decades.

[19]How does the calibrated permanent price impact compare to Gabaix and Koijen (2024), who find that investing \$1 in the stock market increases aggregate value by \$5. For 2025, the SD of monthly net volume is

Panel (b) presents the parameter values that create the perfect match. The fraction of agents who experience a liquidity shock drops from 79% pre-crisis to 54% post-crisis, a drop of 32%. Therefore, there are fewer agents who experience a shock, but the shock they experience increases in size. This is evident from σ_z, which increases from 1.97 to 2.35, an increase of 19%.

The calibration yields the following insights. First, the flow-unncorrelated realized variance, σ_3^2, increases from $(0.49 \times 354) = 173$ to $(0.56 \times 870) = 487$, which is an increase of 182%. All else equal, this implies an increase of price pressure of 182% as per (13). Against this benchmark, an *observed* increase of only 26% suggests that other factors must have caused the market to charge a price pressure far short of 157%. The model achieves this by raising the number of liquidity suppliers, i.e., raising $(1 - \kappa)$, which implies reducing κ (see (13)). This reduced demand and increased supply is able to match the relatively modest increase in price pressure of 26%.

Second, fewer demanders would imply a volume decrease, which is in contrast to the observed increase of 81%. The model achieves this is increase by raising the size of the shock: σ_z increases from 1.97 pre-crisis to 2.35 post-crisis, an increase of 19%. This increase of σ_z seems surprisingly small, because, all else equal, 32% fewer agents (κ) each experiencing a 19% larger shock (σ_z) would imply a volume change of $(1-0.32)\times(1+0.19)-1 = -19\%$. The all-else-equal clause, however, is violated when trading is endogenized, as it is in the model. The increase in liquidity suppliers make the demanders demand more. On the extensive margin, fewer agents are shocked, but on the intensive margin, their shock is larger, *and*

\$524 million. The SD of the monthly, flow-correlated component of long-term SPY price changes is 5.41% times SPY market capitalization (\$617,000 million): \$33,365 million. The ratio of the two SDs is 64, which is a bit more ten times larger than what Gabaix and Koijen find. This could be explained by the net volume measure only capturing about a tenth of total net volume in SPY. The net volume measure used here is based on Arca and Nasdaq volume only and, therefore, misses trading on other-market and off-market trading. This back-of-the-envelope calculation, of course, needs to be taken with a grain of salt, as (i) I document correlation, not causation and (ii) Authorized Participants can arbitrage SPY against the basket of S&P 500 securities.

they demand more liquidity because it is cheaper. These two effects result in an endogenous volume increase that matches the observed increase of 81%.[20]

Third, the (transitory) price pressure that pays liquidity suppliers for their supply, increases pre- to post-crisis. It increases both in an absolute and in a relative sense. Price pressure is meaningfully defined as:

$$\begin{aligned} PricePressure = RealizedVariance - FundamentalVariance = \\ = x_1 - \left(\sigma_{v_2}^2 + \sigma_{v_3}^2\right) = \\ = \left(\sigma_{v_2} + \alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2 - \sigma_{v_2}^2, \end{aligned} \tag{24}$$

where the expression for x_1 is in (20). Appendix G calculates price pressures based on the calibrated values. It shows that, pre-crisis, the price pressure component of monthly realized variance is 5.0 percent squared (out of a total realized variance of 354 percent squared). This increases to 14.9 post-crisis (out of a total of 870), which is an increase of 198%. Note that this increase is larger than the realized variance increase, which is 146%. The larger increase of price pressure is due to both an increase in the per-unit price pressure (+26%), and an increase in volume (+81%).

3 Conclusion

Previous studies document that the variance risk premium is positive in the data, it jumps after a crisis, and it remains elevated for subsequent months. I replicate these findings based on a relatively long time series: 1993-2025. I add to these findings by documenting that trading the index by means of an ETF increases in the post-crisis months. This elevated

[20] Note that the mass of liquidity suppliers, i.e., $(1-\kappa)$, increases by a factor $(1-0.54)/(1-0.79)-1 = 119\%$. Adding this to the earlier calculation yields $((1-0.19)\times(1+1.19)-1 = +77\%$, which is closer to the observed value of +81%. Note that this is a back-of-the-envelope calculation and, therefore, does not perfectly match the endogenous volume increase of 81%.

volume experiences higher costs in the sense that prices overshoot fundamentals to a larger degree in the months after the crisis.

These empirical findings are explained by a model where agents experience liquidity risk. The model has some attractive features. First, all results are relatively straightforward closed-form expressions. Second, the model implies that the variance risk premium is nonnegative, consistent with the data. The reason is that agents use it to hedge against liquidity shocks. The agents who experience a shock and are forced to sell, these agents particularly like to be paid the realized variance. Not only does it compensate for their elevated risk due to the shock, it also compensates for the higher liquidity premium that needs to be paid to hedge part of it. Both are endogenous in the model. Third, the model explains the post-crisis pattern of increased volume, yet liquidity.

These findings provide a novel perspective on VIX dynamics. The model prices variance swaps and, therefore, it prices squared VIX. Regulators should become more vigilant at times when VIX is high relative to real-world volatility. The model tells them that liquidity risk is high in these periods. This is either due to *more* agents experiencing such risk, or *fewer* agents experiencing stronger shocks. Trade data lets them differentiate the two. Calibration of the model to 11 crises experienced in 1993-2025 suggest that it is the latter: The elevated post-crisis VRP and trading pattern in SPY indicate fewer agents experience larger shocks.

Appendix

A VIX squared as an estimate for implied variance

This appendix outlines the argument for why VIX squared approximates implied variance in a model-free way. It illustrates the derivation presented in Carr and Wu (2009), which builds on a result from Carr and Madan (2001). While the derivation in Carr and Wu (2009) is more general in the sense that it includes jumps, the steps in the derivation are the same as the simplified case without jumps presented in this appendix.

Let S_t be the time t spot price of an asset, and F_t its time t futures price of maturity $T > t$. No arbitrage dictates that there exists a risk-neutral probability measure $\mathbb{Q}$ defined on a probability space: $(\Omega, \mathcal{F}, \mathbb{Q})$. F_t solves the following stochastic differential equation:

$$\mathrm{d}F_t = F_t \sigma_{t-} \mathrm{d}B_t, \tag{25}$$

where B_t is a Brownian motion under $\mathbb{Q}$, and the time subscript on σ_{t-} indicates that it is stochastic, but predictable with respect to filtration $\mathcal{F}_t$. Applying Itô's lemma to $\ln(F_t)$ yields:

$$\ln(F_T) = \ln(F_t) + \int_t^T \frac{1}{F_s} \mathrm{d}F_s + \frac{1}{2} \int_t^T \frac{1}{F_s^2} (F_s \sigma_s)^2 \mathrm{d}s = \ln(F_t) + \int_t^T \frac{1}{F_s} \mathrm{d}F_s + \frac{1}{2} \int_t^T \sigma_s^2 \mathrm{d}s. \tag{26}$$

Adding

$$0 = -2 \int_t^T \frac{1}{F_t} \mathrm{d}F_s + 2 \left(\frac{F_T}{F_t} - 1 \right) \tag{27}$$

to (26) and rearranging yields:

$$V_{t,T} := \int_t^T \sigma_s^2 \mathrm{d}s = 2 \left[\frac{F_T}{F_t} - 1 - \ln\left(\frac{F_T}{F_t} \right) \right] + 2 \int_t^T \left(\frac{1}{F_s} - \frac{1}{F_t} \right) \mathrm{d}F_s. \tag{28}$$

Carr and Madan (2001, Appendix 1) use the fundamental theorem of calculus to show the following:

$$\begin{aligned}
f(y) &= f(x) + 1_{y>x} \int_x^y f'(u)\,\mathrm{d}u - 1_{y<x} \int_y^x f'(u)\,\mathrm{d}u = \\
&= f(x) + 1_{y>x} \int_x^y \left[f'(x) + \int_x^u f''(v)\,\mathrm{d}v \right] \mathrm{d}u - 1_{y<x} \int_y^x \left[f'(x) - \int_u^x f''(v)\,\mathrm{d}v \right] \mathrm{d}u = \\
&= f(x) + f'(x)(y-x) + 1_{y>x} \int_x^y \int_v^y f''(v)\,\mathrm{d}u\mathrm{d}v + 1_{y<x} \int_y^x \int_y^v f''(v)\,\mathrm{d}u\mathrm{d}v = \\
&= f(x) + f'(x)(y-x) + 1_{y>x} \int_x^y f''(v)(y-v)\,\mathrm{d}v + 1_{y<x} \int_y^x f''(v)(v-y)\,\mathrm{d}v = \\
&= f(x) + f'(x)(y-x) + \int_x^\infty f''(v)(y-v)^+\,\mathrm{d}v + \int_0^x f''(v)(v-y)^+\,\mathrm{d}v.
\end{aligned} \tag{29}$$

Inserting $f(.) = \ln(.)$, $y = F_T$, and $x = F_t$ into (29) allows us to replace $\ln(F_T)$ in (28) to obtain:

$$V_{t,T} = 2 \left[\int_0^{F_t} \frac{1}{K^2} (K - S_T)^+ \mathrm{d}K + \int_{F_t}^\infty \frac{1}{K^2} (S_T - K)^+ \mathrm{d}K \right] + 2 \int_t^T \left(\frac{1}{F_s} - \frac{1}{F_t} \right) \mathrm{d}F_s. \tag{30}$$

Thus, one can replicate realized variance up to time T by the sum of a static position of $2(\mathrm{d}K)/K^2$ European options on the underlying spot at strike K and maturity T, and the payoff from a dynamic trading strategy holding $2B_s(T)[1/F_s - 1/F_t]$ futures at time s. The options are all out-of-the-money, i.e., put options when $K \leq F_t$ and call options when $F_t > K$.

B Risk premia in CARA-Normal case

Consider a CARA investor with wealth W in cash. His risk aversion coefficient is α. To find the risk premium this investor requires for holding θ units of the risky asset, one needs to compute the certainty equivalent (CE). It is defined as the amount of cash he needs to be offered to become indifferent between holding these assets and receiving CE. Let dividend be stochastic: $v \sim N(0, \sigma_v^2)$.

Asset with linear payoff. The expected utility of letting the investor hold θ units of an asset that pays off v is:

$$
\begin{gathered}
E\left(-\exp\left(-\alpha\left(W+\theta v\right)\right)\right) = \\
\int_v -\frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\alpha\left(W+\theta v\right)\right)\exp\left(-\frac{1}{2}\frac{v^2}{\sigma_v^2}\right)\partial v = \\
-\exp\left(-\alpha W + \frac{1}{2}\alpha^2\theta^2\sigma_v^2\right)\underbrace{\int_v \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{1}{2}\frac{\left(v+\alpha\theta\sigma_v^2\right)^2}{\sigma_v^2}\right)\partial v}_{=1 \text{ (integral over PDF)}} = \\
-\exp\left(-\alpha\left(W - \frac{1}{2}\alpha\theta^2\sigma_v^2\right)\right).
\end{gathered} \tag{31}
$$

The result in (31) shows that the CE for θ units of the risky asset is $\frac{1}{2}\alpha\theta^2\sigma_v^2$. The risk premium per unit of the asset, therefore, is:

$$
\boxed{\text{Risk premium per unit of linear payoff} = \frac{1}{2}\alpha\theta\sigma_v^2.} \tag{32}
$$

Asset with squared payoff. The expected utility of letting the investor hold θ units of an asset that pays off v^2 is:

$$
\begin{gathered}
E\left(-\exp\left(-\alpha\left(W+\theta v^2\right)\right)\right) = \\
\int_v -\frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\alpha\left(W+\theta v^2\right)\right)\exp\left(-\frac{1}{2}\frac{v^2}{\sigma_v^2}\right)\partial v = \\
-\exp\left(-\alpha W\right)\underbrace{\int_v \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{1}{2}\frac{v^2\left(1+2\alpha\theta\sigma_v^2\right)}{\sigma_v^2}\right)\partial v}_{=1 \text{ (integral over PDF)}} = \\
-\exp\left(-\alpha W\right).
\end{gathered} \tag{33}
$$

The result in (33) shows that the CE for θ units of such asset is zero. The risk premium per unit of the asset, therefore, is:

$$
\boxed{\text{Risk premium per unit of squared payoff} = 0.} \tag{34}
$$

C Noisy signal extension

Let agents receive a noisy signal on the dividend in Period 1:

$$s = v + \eta, \quad \eta \sim N\left(0, \sigma_\eta^2\right), \quad \eta \perp v. \tag{35}$$

Total dividend then becomes the sum of two orthogonal Gaussian components:

$$v = v_2 + v_3, \tag{36}$$

where v_2 is the posterior mean, i.e.,

$$v_2 = E\left(v|s\right) = \mu_v + \frac{\mathrm{cov}\left(v, s\right)}{\mathrm{var}\left(s\right)}\left(s - \mu_v\right) = \mu_v + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\left(s - \mu_v\right). \tag{37}$$

The variance of v_2, therefore, is:

$$\sigma_{v_2}^2 = \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\right)^2 \mathrm{var}\left(s - \mu_v\right) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\sigma_v^2. \tag{38}$$

The variance of the orthogonal component v_3, therefore, is:

$$v_3 \sim N\left(0, \sigma_{v_3}^2\right), \quad \sigma_{v_3}^2 = \sigma_v^2 - \sigma_{v_2}^2 = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2}\sigma_v^2. \tag{39}$$

The parameters σ_v^2 and σ_η^2 could then be backed out by (38) and (39).

D Proofs

This appendix provides the proofs for all lemmas and propositions. Important variables or expressions are emphasized by placing them inside a text box.

Proof of Lemma 1

The equilibrium quantities are taken from Section 2 of VW12. In particular, the demand function for the risky asset of a liquidity demander is:

$$\theta_2^d = \frac{v_2 - p_2}{\alpha\sigma_{v_3}^2} - z. \tag{40}$$

For the liquidity supplier it is:

$$\theta_2^s = \frac{v_2 - p_2}{\alpha\sigma_{v_3}^2}. \tag{41}$$

Market clearing implies that the Period 2 equilibrium price is:

$$\boxed{p_2} = v_2 - \alpha\sigma^2_{v_3}\left(\bar{\theta} + \kappa z\right). \tag{42}$$

Therefore:

$$\boxed{\theta^d_2} = \bar{\theta} - (1-\kappa)\, z \tag{43}$$

and

$$\boxed{\theta^s_2} = \bar{\theta} + \kappa z. \tag{44}$$

This implies that (initiator) net volume is (see also VW12, Equation (2.15)):

$$\boxed{q} = \Delta\theta^d_2 = \kappa\left(\theta^d_2 - \bar{\theta}\right) = -\kappa\,(1-\kappa)\, z. \tag{45}$$

Note that this net volume is the exact opposite of net volume of liquidity suppliers:

$$(1-\kappa)\left(\theta^s_2 - \bar{\theta}\right) = (1-\kappa)\,\kappa z. \tag{46}$$

Iliquidity is defined as the price impact of net volume and, therefore, is equal to:

$$\begin{aligned}
\boxed{\lambda_2} &= \frac{\mathrm{Cov}\,(p_2 - p_1, q)}{\mathrm{Var}\,(q)} = \\
&= \frac{\mathrm{Cov}\left(v_2 - \alpha\sigma^2_{v_3}\kappa z, -\kappa\,(1-\kappa)\, z\right)}{\sigma^2_z\kappa^2\,(1-\kappa)^2} = \\
&= \frac{\mathrm{Cov}\left(\frac{-\sigma_{v_2}}{-\kappa(1-\kappa)\sigma_z}\,(-\kappa\,(1-\kappa))\, z, -\kappa\,(1-\kappa)\, z\right)}{\sigma^2_z\kappa^2\,(1-\kappa)^2} + \frac{\mathrm{Cov}\left(-\alpha\sigma^2_{v_3}\kappa z, -\kappa\,(1-\kappa)\, z\right)}{\sigma^2_z\kappa^2\,(1-\kappa)^2} = \\
&= \frac{\sigma_{v_2}}{\kappa\,(1-\kappa)\,\sigma_z} + \frac{\alpha\sigma^2_{v_3}}{1-\kappa}.
\end{aligned} \tag{47}$$

Proof of Lemma 2

This lemma presents the equilibrium price of the risky asset in Period 1, which, in the general case, is:

$$p_1 = \mu_v - \alpha\sigma^2_{v_2}\bar{\theta} - \alpha\sigma^2_{v_3}\bar{\theta} - \frac{\kappa M}{1-\kappa+\kappa M}\Delta_2\bar{\theta}, \tag{48}$$

where

$$\begin{aligned}
M &= \exp\left(\frac{1}{2}\alpha\Delta_3\bar{\theta}^2\right)\sqrt{\frac{1+\Delta_1\kappa^2}{1+\Delta_1\,(1-\kappa)^2 - \alpha^2\sigma^2_{v_3}\sigma^2_z}},\\
\Delta_1 &= \alpha^2\sigma^2_{v_3}\sigma^2_z,\\
\Delta_2 &= \frac{\alpha\sigma^2_{v_3}\Delta_1\kappa}{1+\Delta_1\,(1-\kappa)^2 - \alpha^2\sigma^2_{v_3}\sigma^2_z},
\end{aligned}$$

$$\Delta_3 = \frac{\alpha\sigma_{v3}^2\Delta_1}{1+\Delta_1(1-\kappa)^2 - \alpha^2\sigma_{v3}^2\sigma_z^2}.$$

The Period 1 price of the risky asset is equal to the expected dividend μ_v, minus three terms. The first two terms are the familiar CARA-Normal terms (i.e., $\alpha\sigma_{v_2}^2\bar{\theta}+\alpha\sigma_{v_3}^2\bar{\theta}$). The final term is nonstandard. It captures an additional discount due to excess price volatility caused by Period 2 (transitory) price pressures, which are needed to clear the market. They compensate liquidity suppliers for holding additional risk. Note that this term disappears when there is no liquidity risk and, therefore, no trading, i.e., if $\sigma_z^2 = 0$, then $\Delta_2 = 0$. In a sense, it represents a liquidity *risk* premium. This additional variance is endogenous to the model and is also accounted for when pricing the variance swap.

Proof of Proposition 1

The marginal utility of the swap is:

$$\begin{aligned}\frac{\partial}{\partial\theta_x}\Big(&\kappa E_{t=1}\left[-\exp\left(-\alpha\left(w_d+\theta_x V\right)\right)|d\right]+\\&+(1-\kappa)E_{t=1}\left[-\exp\left(-\alpha\left(w_s+\theta_x V\right)\right)|s\right]\Big)\Big|_{\theta_x=0},\end{aligned} \tag{49}$$

where the subscript d refers to the liquidity-demander type, the subscript s to the liquidity-supplier type, and final-period wealth for type i is w_i. Therefore, the price of the variance swap, expressed in units of the Period 3 consumption good, is:

$$\boxed{y_1} = \kappa\alpha E\left[V\exp\left(-\alpha w_d\right)|d\right]+(1-\kappa)\alpha E\left[V\exp\left(-\alpha w_s\right)|s\right]. \tag{50}$$

The price of the variance swap given in (50) contains expressions in the numerator and in the denominator of the form:

$$E_a\left[\left(a'B_1a+b_2'a+b_3\right)\exp\left(-\alpha\left(a'B_4a+b_5'a+b_6\right)\right)\right], \tag{51}$$

where

$$a = \begin{pmatrix} v_2 \\ z \\ v_3 \end{pmatrix}. \tag{52}$$

The reason for the shape of (51) is:

1. The quadratic expression $(a'B_1a+b_2'a+b_3)$ is the result of realized variance in (15), which is quadratic in price changes. These price changes are, in turn, affine in a.

2. The expression $(a'B_4a+b_5'a+b_6)$ corresponds to the wealth of the agent. It consists of a risk-free position, long or short, due to payment in the Period 2 trading round. It further consists of an exposure to v_3 caused by the post-trade position in the risky asset. This position is affine in a.

The distribution of the random variable a is:

$$a \sim N(\mu, \Sigma), \tag{53}$$

where

$$\mu = \begin{pmatrix} \mu_v & 0 & 0 \end{pmatrix}', \tag{54}$$

$$\Sigma = \begin{pmatrix} \sigma_{v_2}^2 & -\sigma_{v_2}\sigma_z & 0 \\ -\sigma_{v_2}\sigma_z & \sigma_z^2 & 0 \\ 0 & 0 & \sigma_{v_3}^2 \end{pmatrix}. \tag{55}$$

I compute (51) as follows:

$$\begin{aligned}
&\int_{\mathbb{R}^3} (a'B_1a + b_2'a + b_3) \exp(-\alpha(a'B_4a + b_5'a + b_6)) \times \\
&\times \frac{1}{\sqrt{2\kappa|\Sigma|}} \exp\left(-\frac{1}{2}(a-\mu)'\Sigma^{-1}(a-\mu)\right) da = \\
&= \sqrt{\frac{|\Sigma^*|}{|\Sigma|}} \exp\left(-\alpha b_6 + \frac{1}{2}\mu^{*\prime}\Sigma^{*-1}\mu^* - \frac{1}{2}\mu'\Sigma^{-1}\mu\right) \times \\
&\times \int_{\mathbb{R}^3} (a'B_1a + b_2'a + b_3) \frac{1}{\sqrt{2\kappa|\Sigma^*|}} \times \\
&\times \exp\left(-\frac{1}{2}(a-\mu^*)'\Sigma^{*-1}(a-\mu^*)\right) da \\
&= \sqrt{\frac{|\Sigma^*|}{|\Sigma|}} \exp\left(-\alpha b_6 + \frac{1}{2}\mu^{*\prime}\Sigma^{*-1}\mu^* - \frac{1}{2}\mu'\Sigma^{-1}\mu\right) \times \\
&\times E(a^{*\prime}B_1a^* + b_2'a^* + b_3),
\end{aligned} \tag{56}$$

where

$$a^* \sim N(\mu^*, \Sigma^*), \tag{57}$$

and

$$\mu^* = \Sigma^*\left(\Sigma^{-1}\mu - \alpha b_5\right), \tag{58}$$

$$\Sigma^* = \left(\Sigma^{-1} + 2\alpha B_4\right)^{-1}. \tag{59}$$

What remains is to compute the expectation in (56):

$$E(a^{*\prime}B_1a^* + b_2'a^* + b_3) = E\left(\sum_i\sum_j (B_1)_{ij}\, a_i^* a_j^*\right) + b_2'\mu^* + b_3 =$$

$$
\begin{aligned}
&= \sum_i \sum_j (B_1)_{ij} \left(\sigma^*_{ij} + \mu_i \mu_j\right) + b_2' \mu^* + b_3 = \\
&= \sum_i \sum_j (B_1)_{ij} \sigma^*_{ji} + \mu^{*\prime} B_1 \mu^* + b_2' \mu^* + b_3 = \\
&= \sum_i (B_1 \Sigma^*)_{i,i} + \mu^{*\prime} B_1 \mu^* + b_2' \mu^* + b_3 = \\
&= tr\,(B_1 \Sigma^*) + \mu^{*\prime} B_1 \mu^* + b_2' \mu^* + b_3. \qquad (60)
\end{aligned}
$$

Putting it all together yields the following expressions for the numerator and the denominator of (50):

$$
\begin{aligned}
\boxed{y_1} &= E\Big[(a' B_1 a + b_2' a + b_3) \times \\
&\times \exp\left(-\alpha \left(a' B_4 a + b_5' a + b_6\right)\right) \Big] = \\
&= \sqrt{\frac{|\Sigma^*|}{|\Sigma|}} \exp\left(-\alpha b_6 + \frac{1}{2} \mu^{*\prime} \Sigma^{*-1} \mu^* - \frac{1}{2} \mu' \Sigma^{-1} \mu\right) \times \\
&\times \left(tr\,(B_1 \Sigma^*) + \mu^{*\prime} B_1 \mu^* + b_2' \mu^* + b_3\right), \qquad (61)
\end{aligned}
$$

with

$$
\mu^* = \Sigma^* \left(\Sigma^{-1} \mu - \alpha b_5\right) \quad \text{and} \quad \Sigma^* = \left(\Sigma^{-1} + 2\alpha B_4\right)^{-1}. \qquad (62)
$$

Expressions for B_1, b_2, b_3, B_4, b_5, and b_6 in (51)**.** Now that the equilibrium price of the variance swap has been derived analytically, what remains is some administrative work. The expressions for B_1, b_2, b_3, B_4, b_5, and b_6 in (61) need to be written down explicitly. Since all these quantities are affine in the state variables, I use subscripts to distinguish between the coefficient of the state variable a, and the constant. For example, the position in the risky asset of the liquidity demander after trading in Period 1 is:

$$
\theta_2^d = \bar{\theta} - (1 - \kappa)\, z = \theta_{1,a}^d a + \theta_{1,c}^d, \qquad (63)
$$

where

$$
\begin{aligned}
\theta_{1,a}^d &= \begin{pmatrix} 0 & -(1-\kappa) & 0 \end{pmatrix}', \qquad (64) \\
\theta_{1,c}^d &= \bar{\theta}. \qquad (65)
\end{aligned}
$$

With this notation, the constants for the liquidity demander in terms of the equilibrium quantities can be expressed as follows:

$$
\begin{aligned}
B_1 &= (p_{1,a} - p_{0,a})(p_{1,a} - p_{0,a})' + (p_{2,a} - p_{1,a})(p_{2,a} - p_{1,a})', \qquad (66) \\
b_2 &= 2\,(p_{1,c} - p_{0,c})(p_{1,a} - p_{0,a}) + 2\,(p_{2,c} - p_{1,c})(p_{2,a} - p_{1,a}), \qquad (67) \\
b_3 &= (p_{1,c} - p_{0,c})^2 + (p_{2,c} - p_{1,c})^2, \qquad (68)
\end{aligned}
$$

$$B_4 = \underbrace{-\frac{1}{2}\left(\Delta\theta^d_{1,a} p'_{1,a} + p_{1,a}\Delta\theta^{d\prime}_{1,a}\right)}_{\text{Payment made}} +$$

$$+ \underbrace{\frac{1}{2}\left(\left(\theta^d_{1,a} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right)\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\left(\theta^{d\prime}_{1,a} + \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}\right)\right)}_{\text{Exposure}}, \tag{69}$$

$$b_5 = -\Delta\theta^d_{1,c} p_{1,a} - p_{1,c}\Delta\theta^d_{1,a} + \theta^d_{1,c}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{70}$$

$$b_6 = \theta^d_{1,c} p_{1,c}. \tag{71}$$

For a liquidity supplier, $B1$, $b2$, and b_3 are the same, but the remaining ones need to be changed to:

$$B_4 = \underbrace{\frac{1}{2}\left(\Delta\theta^s_{1,a} p'_{1,a} + p_{1,a}\Delta\theta^{s\prime}_{1,a}\right)}_{\text{Payment received}} + \underbrace{\frac{1}{2}\left(\theta^s_{1,a}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\theta^{s\prime}_{1,a}\right)}_{\text{Exposure}}, \tag{72}$$

$$b_5 = \Delta\theta^s_{1,c} p_{1,a} + p_{1,c}\Delta\theta^s_{1,a} + \theta^s_{1,c}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{73}$$

$$b_6 = \theta^s_{1,c} p_{1,c}. \tag{74}$$

Special case: $\bar{\theta} = \mu_v = 0$.

To develop further insight, assume that $\bar{\theta} = \mu_v = 0$ as per Assumption 1. That is, both the supply of the risky asset and the mean dividend are set to zero, respectively. The expressions in this subsection have been double-checked by means of a Mathematica notebook that is available at `https://bit.ly/3CM8lDZ`.

Liquidity demander. For the liquidity demander, marginal utility for the component of realized variance associated with the price change from Period 1 to 2 is equal to (61) with:

$$B_1 = \begin{pmatrix} -\left(\frac{\sigma_{v_2}}{\sigma_z} + \alpha\sigma^2_{v_3}\kappa\right) \\ 0 \end{pmatrix}\begin{pmatrix} -\left(\frac{\sigma_{v_2}}{\sigma_z} + \alpha\sigma^2_{v_3}\kappa\right) & 0 \end{pmatrix} +$$

$$+ \begin{pmatrix} \alpha\sigma^2_{v_3}\kappa \\ 1 \end{pmatrix}\begin{pmatrix} \alpha\sigma^2_{v_3}\kappa & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2_{v_2}}{\sigma^2_z} + 2\alpha\sigma^2_{v_3}\kappa\frac{\sigma_{v_2}}{\sigma_z} + 2\alpha^2\sigma^4_{v_3}\kappa^2 & \alpha\sigma^2_{v_3}\kappa \\ \alpha\sigma^2_{v_3}\kappa & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} \left(\frac{\sigma_{v_2}}{\sigma_z} + \alpha\sigma^2_{v_3}\kappa\right)^2 + \left(\alpha\sigma^2_{v_3}\kappa\right)^2 & \alpha\sigma^2_{v_3}\kappa \\ \alpha\sigma^2_{v_3}\kappa & 1 \end{pmatrix}, \tag{75}$$

$$
\begin{aligned}
b_2 &= \begin{pmatrix} 0 & 0 \end{pmatrix}', && (76)\\
b_3 &= 0, && (77)\\
B_4 &= \frac{1}{2}\begin{pmatrix} 1-\kappa \\ 0 \end{pmatrix}\begin{pmatrix} -\alpha\sigma_{v3}^2\kappa & 0 \end{pmatrix} + \frac{1}{2}\left(\begin{pmatrix} 1-\kappa \\ 0 \end{pmatrix}\begin{pmatrix} -\alpha\sigma_{v3}^2\kappa & 0 \end{pmatrix}\right)' + \\
&\quad + \frac{1}{2}\begin{pmatrix} -(1-\kappa)+1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix} + \frac{1}{2}\left(\begin{pmatrix} -(1-\kappa)+1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}\right)' = \\
&= \begin{pmatrix} -\alpha\sigma_{v3}^2\kappa(1-\kappa) & \frac{\kappa}{2} \\ \frac{\kappa}{2} & 0 \end{pmatrix}, && (78)\\
b_5 &= \begin{pmatrix} 0 & 0 \end{pmatrix}', && (79)\\
b_6 &= 0. && (80)
\end{aligned}
$$

Note that B_1 in (75) features *positive* off-diagonal elements. The reason is that, although price pressures mean-revert and thus do not create a wealth loss to the aggregate investor, they *do* add to realized variance. As the payoff to the variance swap is this realized variance, it adds to total variance and, therefore, the off-diagonals are positive.

The values in (75) through (80) for this special case environment imply (as per (62)):

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_{v3}^2 \end{pmatrix}, && (81)\\
\mu &= \begin{pmatrix} 0 & 0 \end{pmatrix}', && (82)\\
\Sigma^* &= \left(\Sigma^{-1} + 2\alpha B_4\right)^{-1} \\
&= \begin{pmatrix} \frac{1}{\sigma_z^2} - 2\alpha^2\sigma_{v3}^2\kappa(1-\kappa) & \alpha\kappa \\ \alpha\kappa & \frac{1}{\sigma_{v3}^2} \end{pmatrix}^{-1}, \\
&= \frac{1}{1-\alpha^2\sigma_{v3}^2\sigma_z^2\kappa(2-\kappa)}\begin{pmatrix} \sigma_z^2 & -\alpha\sigma_{v3}^2\sigma_z^2\kappa \\ -\alpha\sigma_{v3}^2\sigma_z^2\kappa & \sigma_{v3}^2\left(1-2\alpha^2\sigma_{v3}^2\sigma_z^2\kappa(1-\kappa)\right) \end{pmatrix}, && (83)\\
\mu^* &= \Sigma^*\left(\Sigma^{-1}\mu - \alpha b_5\right) \\
&= \begin{pmatrix} 0 & 0 \end{pmatrix}'. && (84)
\end{aligned}
$$

Note that to value the variance swap, the real-world probability measure, often referred to as the $\mathbb{P}$-measure, is replaced by a risk-neutral measure, often known as the $\mathbb{Q}$-measure. The latter features a variance of Σ^* and, comparing its expression in (83), to the expression of Σ in (81), shows that the $\mathbb{Q}$-measure increases variance for both z and v_3. The off-diagonal is negative, because price pressures are transitory. These price pressures, therefore, add to realized variance (as per our discussion about B_1 above), but they do not affect wealth in Period 3 for the aggregate investor.

Therefore, the marginal utility of the risky asset for d is:

$$\boxed{y_1^d} = E\left[\left(a'B_1a + b_2'a + b_3\right)\exp\left(-\alpha\left(a'B_4a + b_5'a + b_6\right)\right)\right] =$$
$$= \sqrt{\frac{|\Sigma^*|}{|\Sigma|}} tr\left(B_1\Sigma^*\right) =$$
$$\boxed{= \frac{1}{1-\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa\left(2-\kappa\right)}\left(\frac{\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2}{1-\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa\left(2-\kappa\right)}+\sigma_{v_3}^2\right),} \tag{85}$$

because[21]

$$|\Sigma^*| = \frac{\sigma_{v_3}^2\sigma_z^2}{1-\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa\left(2-\kappa\right)}, \quad |\Sigma| = \sigma_{v_3}^2\sigma_z^2,$$
$$tr\left(B_1\Sigma^*\right) = \frac{\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2}{1-\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa\left(2-\kappa\right)}+\sigma_{v_3}^2. \tag{86}$$

Liquidity supplier. For the liquidity supplier, all coefficients are the same except for:

$$B_4 = \frac{1}{2}\begin{pmatrix}-\kappa\\0\end{pmatrix}\begin{pmatrix}-\alpha\sigma_{v_3}^2\kappa & 0\end{pmatrix} + \frac{1}{2}\left(\begin{pmatrix}-\kappa\\0\end{pmatrix}\begin{pmatrix}-\alpha\sigma_{v_3}^2\kappa & 0\end{pmatrix}\right)'$$
$$+\frac{1}{2}\begin{pmatrix}\kappa\\0\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix} + \frac{1}{2}\left(\begin{pmatrix}\kappa\\0\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}\right)' =$$
$$= \begin{pmatrix}\alpha\kappa^2\sigma_{v_3}^2 & \frac{\kappa}{2}\\ \frac{\kappa}{2} & 0\end{pmatrix}. \tag{87}$$

These values imply that the following objects change relative to the liquidity demander case:

$$\Sigma^* = \begin{pmatrix}\frac{1}{\sigma_z^2}+2\kappa^2\alpha^2\sigma_{v_3}^2 & \alpha\kappa\\ \alpha\kappa & \frac{1}{\sigma_{v_3}^2}\end{pmatrix}^{-1}$$
$$= \frac{1}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}\begin{pmatrix}\sigma_z^2 & -\alpha\sigma_{v_3}^2\sigma_z^2\kappa\\ -\alpha\sigma_{v_3}^2\sigma_z^2\kappa & \sigma_{v_3}^2\left(1+2\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2\right)\end{pmatrix}, \tag{88}$$

therefore,

$$|\Sigma^*| = \frac{\sigma_{v_3}^2\sigma_z^2}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}, \quad tr\left(B_1\Sigma^*\right) = \frac{\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z^2\kappa\right)^2}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}+\sigma_{v_3}^2. \tag{89}$$

[21] Note that the outcome is well defined given the condition in (10).

The marginal utility of a variance swap to the liquidity supplier, therefore, is:

$$\boxed{y_1^s = \sqrt{\frac{|\Sigma^*|}{|\Sigma|}} tr\left(B_1\Sigma^*\right) = \frac{1}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}\left(\frac{\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}+\sigma_{v_3}^2\right).} \tag{90}$$

So, the marginal utility of the variance swap in Period 1 is obtained by taking the expected marginal utility, i.e., multiplying (85) and (90) by κ and $(1-\kappa)$, respectively:

$$\begin{aligned}
\boxed{y_1} &= \kappa y_1^d + (1-\kappa)\, y_1^s = \\
&= \kappa \underbrace{\left(A^2\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2 + A\sigma_{v_3}^2\right)}_{\text{Valuation liquidity demander}} + \\
&+ (1-\kappa)\underbrace{\left(B^2\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2 + B\sigma_{v_3}^2\right)}_{\text{Valuation liquidity supplier}} = \\
&= \underbrace{\left(\kappa A^2 + (1-\kappa)\, B^2\right)}_{\text{Inflator corr. return}} \times \underbrace{\left(\sigma_{v_2}+\alpha\sigma_{v_3}^2\sigma_z\kappa\right)^2}_{\text{Variance corr. return}} + \\
&+ \underbrace{\left(\kappa A + (1-\kappa)\, B\right)}_{\text{Inflator noncorr. return}} \times \underbrace{\sigma_{v_3}^2}_{\text{Variance noncorr. return}},
\end{aligned} \tag{91}$$

with

$$A = \frac{1}{1-\alpha^2\sigma_{v_3}^2\sigma_z^2\left(2-\kappa\right)\kappa} \tag{92}$$

and

$$B = \frac{1}{1+\alpha^2\sigma_{v_3}^2\sigma_z^2\kappa^2}. \tag{93}$$

The proof for the final statement that both inflators are larger than one follows from a result in the proof of Proposition 2. When proving that y_1 increases in κ, it is shown that both inflators increase in κ. This result, along with the observation that both inflators are one for $\kappa = 0$, shows that both inflators are at least one. What remains to show is:

$$\kappa A^2 + (1-\kappa)\, B^2 \geq \kappa A + (1-\kappa)\, B. \tag{94}$$

Applying Jensen's inequality to the convex function $f(x) = x^2$ with the stochastic variable:

$$X = \begin{cases} A \text{ with probability } \kappa, \\ B \text{ with probability } 1-\kappa, \end{cases} \tag{95}$$

yields

$$\kappa A^2 + (1-\kappa)\, B^2 \geq \kappa A + (1-\kappa)\, B. \tag{96}$$

This completes the proof.

Proof of Proposition 2

The algebra in this subsection has been double-checked by means of a Mathematica notebook that is available at `https://bit.ly/3CM81DZ`. The engine of the monotonicity proof is the following lemma.

Lemma 3 *Let*

$$f_{c,n}(\kappa) = \frac{1}{(1-(2-\kappa)\kappa c)^n} - \frac{1}{(1+c\kappa^2)^n}, \tag{97}$$

where

$$c \in [0,1), \quad n \in \{1,2,3,\ldots\}. \tag{98}$$

Then $f_{c,n}(0) = 1$ and $f_{c,n}(\kappa)$ increases monotonically in the interval $\kappa \in [0,1]$. Therefore, $f_{c,n}(\kappa)$ is positive in this interval.

To prove Lemma 3, first define

$$g_c(\kappa) = \frac{1}{1-(2-\kappa)\kappa c} - \frac{1}{1+c\kappa^2}, \tag{99}$$

then

$$g_c'(\kappa) = \frac{2c\left(c^2(4-3\kappa)\kappa^3 - 2c\kappa^2 + 1\right)}{(1-c(2-\kappa)\kappa)^2\left(c\kappa^2+1\right)^2}, \tag{100}$$

and the sign of this derivative, therefore, depends on the sign of

$$h_c(\kappa) = c^2(4-3\kappa)\kappa^3 - 2c\kappa^2 + 1. \tag{101}$$

Now, from

$$h_c'(\kappa) = -4c\kappa\left(3c(\kappa-1)\kappa + 1\right) < 0, \tag{102}$$

$c \in [0,1)$, $\kappa(1-\kappa) \leq 1/4$ for $\kappa \in [0,1]$, $h_c(0) = 1$, $h_c(1) = 0$, and $h_c(\kappa)$ being a continuous differentiable function, it follows that $h_c(\kappa)$ is indeed strictly positive on the domain $\kappa \in [0,1)$, based on the Intermediate Value Theorem applied to $\kappa \in [0, 1+\varepsilon]$ with $\varepsilon > 0$. This implies that $g_c'(\kappa) > 0$ and $g_c(\kappa)$, therefore, increases monotonically in κ.

With the proof that $g_c(\kappa)$ increases monotonically in κ, what remains to be proven is that this implies that $f_{c,n}(\kappa)$ increases monotonically in κ. This follows from:

$$\begin{aligned} f_{c,n}'(\kappa) = n&\left(\frac{1}{(1-(2-\kappa)\kappa c)}\right)^{n-1}\frac{\partial}{\partial\kappa}\left(\frac{1}{(1-(2-\kappa)\kappa c)}\right) + \\ -n&\left(\frac{1}{1+c\kappa^2}\right)^{n-1}\frac{\partial}{\partial\kappa}\left(\frac{1}{1+c\kappa^2}\right) \\ \geq n&\left(\frac{1}{(1-(2-\kappa)\kappa c)}\right)^{n-1}\frac{\partial}{\partial\kappa}\left(\frac{1}{(1-(2-\kappa)\kappa c)}\right) + \end{aligned} \tag{103}$$

$$- n\left(\frac{1}{(1-(2-\kappa)\,\kappa c)}\right)^{n-1}\frac{\partial}{\partial\kappa}\left(\frac{1}{1+c\kappa^2}\right) \tag{104}$$

$$= n\left(\frac{1}{(1-(2-\kappa)\,\kappa c)}\right)^{n-1} g_c'(\kappa)\,. \tag{105}$$

As $g_c'(\kappa) > 0$, it follows that $f_c'(\kappa) > 0$, which proves Lemma 3.

To prove a monotonic relationship for the VRP, i.e., $y_1 - x_1$, requires proving it for the two inflators in (16). This is what is done in the remainder of the proof. Let:

$$c := \alpha^2\sigma_{v3}^2\sigma_z^2, \tag{106}$$

which, by condition (10), is in $[0, 1)$, because $\sigma_{v3}^2 \leq \sigma_v^2$.

Proof that $(y_1 - x_1)$ increases monotonically in κ. Let us first the derivative of the flow-correlated inflator in (16):

$$\frac{\partial}{\partial\kappa}\left(\kappa A^2 + (1-\kappa)\,B^2\right) = f_{c,1}(\kappa) + 4\,(1-\kappa)\,\kappa c \times f_{c,3}(\kappa) > 0. \tag{107}$$

The derivative for the flow-uncorrelated inflator in (16) is:

$$\frac{\partial}{\partial\kappa}\left(\kappa A + (1-\kappa)\,B\right) = f_{c,2}(\kappa) + 2\,(1-\kappa)\,\kappa c \times f_{c,2}(\kappa) > 0. \tag{108}$$

Since both variance inflators increase monotonically in κ, y_1 increases monotonically in κ.

Proof that $(y_1 - x_1)$ increases monotonically in α. This proof follows the same steps as the previous one:

$$\begin{aligned}\frac{\partial}{\partial\alpha}\left(\kappa A^2\,(1-\kappa)\,B^2\right) &= \frac{4\kappa^2 c}{\alpha}\left(\frac{2-\kappa}{(1-(2-\kappa)\kappa c)^3} - \frac{1-\kappa}{(1+\kappa^2 c)^3}\right)\\ &\geq \left(\frac{4\kappa^2\,(1-\kappa)\,c}{\alpha}\right) f_{c,3}(\kappa) > 0\end{aligned} \tag{109}$$

and

$$\begin{aligned}\frac{\partial}{\partial\alpha}\left(\kappa A\,(1-\kappa)\,B\right) &= \frac{2\kappa^2 c}{\alpha}\left(\frac{2-\kappa}{(1-(2-\kappa)\kappa c)^2} - \frac{1-\kappa}{(1+\kappa^2 c)^2}\right)\\ &\geq \left(\frac{2\kappa^2\,(1-\kappa)\,c}{\alpha}\right) f_{c,2}(\kappa) > 0.\end{aligned} \tag{110}$$

Therefore both variance inflators increase monotonically in α and thus y_1 increases monotonically in α.

Proof that $(y_1 - x_1)$ increases monotonically in σ_z. This proof follows the same steps as the previous one:

$$\begin{aligned}\frac{\partial}{\partial \sigma_z}\left(\kappa A^2 (1-\kappa) B^2\right) &= \frac{4\kappa^2 c}{\sigma_z}\left(\frac{2-\kappa}{(1-(2-\kappa)\kappa c)^3} - \frac{1-\kappa}{(1+\kappa^2 c)^3}\right) \\ &\geq \left(\frac{4\kappa^2 (1-\kappa) c}{\sigma_z}\right) f_{c,3}(\kappa) > 0 \end{aligned} \tag{111}$$

and

$$\begin{aligned}\frac{\partial}{\partial \sigma_z}\left(\kappa A (1-\kappa) B\right) &= \frac{2\kappa^2 c}{\sigma_z}\left(\frac{2-\kappa}{(1-(2-\kappa)\kappa c)^2} - \frac{1-\kappa}{(1+\kappa^2 c)^2}\right) \\ &\geq \left(\frac{2\kappa^2 (1-\kappa) c}{\sigma_z}\right) f_{c,2}(\kappa) > 0. \end{aligned} \tag{112}$$

Therefore both variance inflators increase monotonically in σ_z and thus y_1 increases monotonically in σ_z.

E Sample construction

This appendix presents details on how the sample is constructed from the raw data sources. The names of these variables in the raw daily files are in typewriter font. The following variables are taken from the *WRDS Intraday Indicators* database, which is available in WRDS TAQ:

- Daily net volume in SPY, which is defined as the sum of Nasdaq net volume plus NYSE-Arca net volume. Nasdaq and NYSE-Arca run limit order markets and trades can, therefore, be signed reliably. The variables used to construct net volume are: `BuyVol_LR_Arca`, `SellVol_LR_Arca`, `BuyVol_LR_Nasd`, and `SellVol_LR_Nasd`.
- The effective spread and realized spread for SPY are share-weighted and based on `EffectiveSpread_Percent_SW` and `PercentRealizedSpread_LR_SW`, respectively.
- Daily volume in SPY, which includes all trade volume during market hours, is equal to `total_vol_m` (i.e., more than Nasdaq and NYSE-Arca).

The S&P index level, SPX, is taken from daily WRDS CRSP `spindx`. VIX is obtained from the historical daily VIX file available at the CBOE website (`Close`).

Proxy construction. For each month in the sample, proxies are constructed for the variance of v_2 and v_3. To remove any transitory effects, these proxies are constructed as follows. The proxy for the annualized variance of v_2 is:

$$\hat{\sigma}^2_{v_2} = 250 \times \text{cov}(\mathit{DailyNetVolume}_t, \Delta \log \mathit{SPX}_{t-1} + \Delta \log \mathit{SPX}_t + \Delta \log \mathit{SPX}_{t+1}), \tag{113}$$

where yesterday's return and tomorrow's return are added to today's return to remove any potential contamination by transitory "noise" in end-of-day index levels. Assume that the log of each day's closing index value is contaminated by orthogonal, serially uncorrelated transitory noise ε_t. Let w_t denote the innovation in the index, i.e., the random-walk increment. Then:

$$\Delta \log SPX_t = \log SPX_t - \log SPX_{t-1} = w_t + \varepsilon_t - \varepsilon_{t-1}, \tag{114}$$

and, therefore,

$$\Delta \log SPX_{t-1} + \Delta \log SPX_t + \Delta \log SPX_{t+1} = w_{t-1} + w_t + w_{t+1} - \varepsilon_{t-2} + \varepsilon_{t+1}. \tag{115}$$

Therefore, the covariance in (113) measures the covariance of today's net volume with today's innovation in the index w_t, *even* if net volume correlates with yesterday's or today's noise in the index.

Following the same logic, the proxy for the annualized variance of v is:

$$\hat{\sigma}_v^2 = 250 \times \text{cov}(\Delta \log SPX_t, \Delta \log SPX_{t-1} + \Delta \log SPX_t + \Delta \log SPX_{t+1}). \tag{116}$$

Therefore, the proxy for annualized variance of v_3 is:

$$\hat{\sigma}_{v_3}^2 = \hat{\sigma}_v^2 - \hat{\sigma}_{v_2}^2. \tag{117}$$

F Details calibration procedure

The calibration involves the following steps. Fix an equidistant grid for the pre-crisis value of κ. Iterate over this grid. For each value of κ, do the following:

1. Compute the liquidity shock size σ_z^2 that delivers the observed pre-crisis VRP using (16).

2. Use the expression for the (transitory) price pressure in (13), the pre-crisis κ, the observed pre- and post flow-uncorrelated return variance ($\sigma_{v_3}^2$), and the observed relative change in price pressure to compute the implied post-crisis κ.

3. Use the expression for net volume in (12) to compute expected volume, which is the expected absolute value of net volume. Expected volume, therefore, equals $\kappa(1-\kappa)\sigma_z\sqrt{\frac{2}{\kappa}}$, because z is Gaussian. The relative change in volume is used to compute post-crisis σ_z^2.

4. Use these post-crisis values for κ and σ_z^2 and the observed post-crisis flow-correlated and the flow-uncorrelated price variance to compute the implied post-crisis VRP.

After all these computations are done, pick the pre-crisis value of κ that delivers a post-crisis VRP that is closest to the observed post-crisis VRP.

G Price pressure calculations

This section presents the calculations of monthly price pressure variance, both pre- and post-crisis.

Pre-crisis. The pre-crisis calculation is as follows. The monthly flow-uncorrelated realized variance is $\sigma_{v_3}^2 = 0.49 \times 354/12 = 14.5$ percent squared. Therefore, $\alpha\sigma_{v_3}^2\sigma_z\kappa = 3 \times 0.00145 \times 1.97 \times 0.79 = 0.7$ percent (see (20)). This implies $\sigma_{v_2} = \sqrt{0.51 \times 354/12} - 0.7 = 3.2$ percent. Price pressure as defined in (24), therefore, is: $(3.2 + 0.7)^2 - 3.2^2 = 5.0$ percent squared.

Post-crisis. Following the same steps for the post-crisis period yields the following numbers. $\sigma_{v_3}^2 = 0.56 \times 870/12 = 40.6$ percent squared. Therefore, $\alpha\sigma_{v_3}^2\sigma_z\kappa = 3 \times 0.00406 \times 2.35 \times 0.54 = 1.5$ percent. This implies $\sigma_{v_2} = \sqrt{0.44 \times 870/12} - 1.5 = 4.1$ percent. Price pressure is: $(4.2 + 1.5)^2 - 4.2^2 = 14.9$ percent squared.

References

Bakshi, Gurdip and Nikunj Kapadia (2003). Delta-Hedged Gains and the Negative Market Volatility Risk Premium. *Review of Financial Studies* 16, pp. 527–566. DOI: `10.1093/rfs/hhg002`.

Bekaert, Geert, Eric Engstrom and Andrey Ermolov (2023). The Variance Risk Premium in Equilibrium Models. *Review of Finance* 27, pp. 1977–2014. DOI: `10.1093/rof/rfad005`.

Bekaert, Geert and Marie Hoerova (2014). The VIX, the Variance Premium and Stock Market Volatility. *Journal of Econometrics* 183, pp. 181–192.

Bollerslev, Tim, George Tauchen and Hao Zhou (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22, pp. 4464–4465.

Carr, Peter and Dilip Madan (2001). Optimal Positioning in Derivative Securities. *Quantitative Finance* 1, pp. 19–37. DOI: `10.1080/713665549`.

Carr, Peter and Liuren Wu (2006). A Tale of Two Indices. *Journal of Derivatives* 13, pp. 13–29.

— (2009). Variance Risk Premiums. *Review of Financial Studies* 22, pp. 1311–1341.

Cheng, Ing-Haw (2019). The VIX Premium. *Review of Financial Studies* 32, pp. 180–227.

Choi, Hoyong, Philippe Mueller and Andrea Vedolin (2017). Bond Variance Risk Premiums. *Review of Finance* 21, pp. 987–1022.

Drechsler, Itamar, Alan Moreira and Alexi Savov (2021). *Liquidity and Volatility*. Manuscript. Wharton.

Dumas, Bernard (1998). Two-Person Dynamic Equilibrium in the Capital Market. *Review of Financial Studies* 2, pp. 157–188. DOI: `10.1093/rfs/2.2.157`.

Evans, Martin D.D. and Richard K. Lyons (2002). Order Flow and Exchange Rate Dynamics. *Journal of Political Economy* 110, pp. 170–180.

Fournier, Mathieu and Kris Jacobs (2020). A Tractable Framework for Option Pricing with Dynamic Market Maker Inventory and Wealth. *Journal of Financial and Quantitative Analysis* 55, pp. 1117–1162. DOI: 10.1017/S0022109019000462.

Gabaix, Xavier and Ralph S.J. Koijen (2024). *In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis*. Manuscript. University of Chicago.

Gârleanu, Nicolae, Lasse H. Pedersen and Allen M. Poteshman (2009). Demand-Based Option Pricing. *Review of Financial Studies* 22, pp. 4259–4299. DOI: 10.1093/rfs/hhp005.

Glosten, Lawrence and Paul Milgrom (1985). Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Agents. *Journal of Financial Economics* 14, pp. 71–100.

Grossman, Sanford J. and Merton H. Miller (1988). Liquidity and Market Structure. *Journal of Finance* 43, pp. 617–633.

Holden, Craig W. and Stacey Jacobsen (2014). Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions. *Journal of Finance* 69, pp. 1747–1785. DOI: 10.1111/jofi.12127.

Konstantinidi, Eirini and George Skiadopoulos (2016). How Does the Market Variance Risk Premium Vary Over Time? Evidence From S&P 500 Variance Swap Investment Returns. *Journal of Banking and Finance* 62, pp. 62–75.

Lochstoer, Lars A. and Tyler Muir (2022). Volatility Expectations. *Journal of Finance* 77, pp. 1055–1096.

Martin, Ian W.R. and Dimitris Papadimitriou (2022). Sentiment and Speculation in a Market with Heterogeneous Beliefs. *American Economic Review* 112, pp. 2465–2517.

Todorov, Viktor (2010). Variance Risk-Premium Dynamics: The Role of Jumps. *Review of Financial Studies* 23, pp. 345–383.

Vayanos, Dimitri (2004). *Flight to Quality, Flight to Liquidity, and the Pricing of Risk*. NBER Working Paper Series #10327. London School of Economics.

Vayanos, Dimitri and Jiang Wang (2012). Liquidity and Asset Returns Under Asymmetric Information and Imperfect Competition. *Review of Financial Studies* 25, pp. 1339–1365.

Table 1: Summary statistics. This table presents summary statistics for the monthly data sample that runs from February 1993 through August 2025.

	Mean	SD	Auto	Skew	Kurt	N
Realized var SPX EOM ($\%^2$, ann.)	342	692	0.48	8.15	84.96	391
Implied var SPX BOM ($\%^2$, ann.)	444	413	0.76	3.45	17.26	391
Realized var risk premium EOM ($\%^2$, ann.)	101	560	0.14	-8.81	104.12	391

Table 2: Model calibration to match crisis patterns. This table calibrates the model to pre- and post-crisis trading. The patterns are based on 11 crisis periods from 1993 through 2025. The pre-crisis period is the three-month period leading up to the crisis, and the post-crisis period is the three-month period following a crisis (see Figure 6).

	Data	Model
Panel (a): Matched moments		
Pre-crisis variance risk premium (%^{2}, ann.)	19	19
Post-crisis variance risk premium (%^{2}, ann.)	319	319
Change in realized spread pre- to post-crisis	+26%	+26%
Change in share volume pre- to post-crisis	+81%	+81%
Panel (b): Calibrated parameters		
Pre-crisis κ (fraction of agents experiencing liquidity shock)		0.79
Post-crisis κ		0.54
Change in κ pre- to post-crisis		-32%
Pre-crisis σ_z (standard deviation liquidity shock)		1.97
Post-crisis σ_z		2.35
Change in σ_z pre- to post-crisis		+19%
Panel (c): Further input calibration		
Pre-crisis realized var SPX EOM (%^{2}, ann.)		354
Pre-crisis implied var SPX BOM (%^{2}, ann.)		373
Pre-crisis IV to RV ratio		1.05
Pre-crisis fraction of RV that is flow-correlated		0.51
Post-crisis realized var SPX EOM (%^{2}, ann.)		870
Post-crisis implied var SPX BOM (%^{2}, ann.)		1189
Post-crisis IV to RV ratio		1.37
Post-crisis fraction of RV that is flow-correlated		0.44
Risk-aversion parameter α		3

t=1: Variance swap priced

- Agents homogenous *ex-ante*.
- Risky asset price denoted p_1.
- Expected realized variance: $x_1 = E_{t=1}(V)$, where $V = \sum_{\tau=2}^{3}(p_\tau - p_{\tau-1})^2$.
- Price variance swap: $y_1 = E_{t=1}\ (V \times m_3)$, where m_3 is the agent's marginal utility at $t = 3$.

t=2: Dividend signal + liquidity shocks

- Dividend revealed partially: $v_2 = E_{t=2}\,(v)$.
- Agent learns his type: Either he will experience a liquidity shock and becomes a liquidity demander, or not.
- Liquidity demanders trade with others, who *de facto* become suppliers.
- Outcome is vector: (price, net volume, illiquidity) or, shorter, $(p, q = \Delta\vartheta^d, \lambda)$.

t=3: Dividend realized

- Dividend v revealed in full, i.e., agents learn remaining component: $v_3 = v - v_2$.
- Risky asset price equal to dividend: $p_3 = v_2 + v_3$.
- Long side of variance swap receives V.

$$\boxed{(p_1, x_1, y_1)} \rightarrow \boxed{(p_2, q_2, \lambda_2)} \rightarrow \boxed{(p_3)}$$

Figure 1: Model summary. This schematic summarizes the model by describing the events in the three periods. The flow diagram at the bottom shows the key variables determined in these periods.

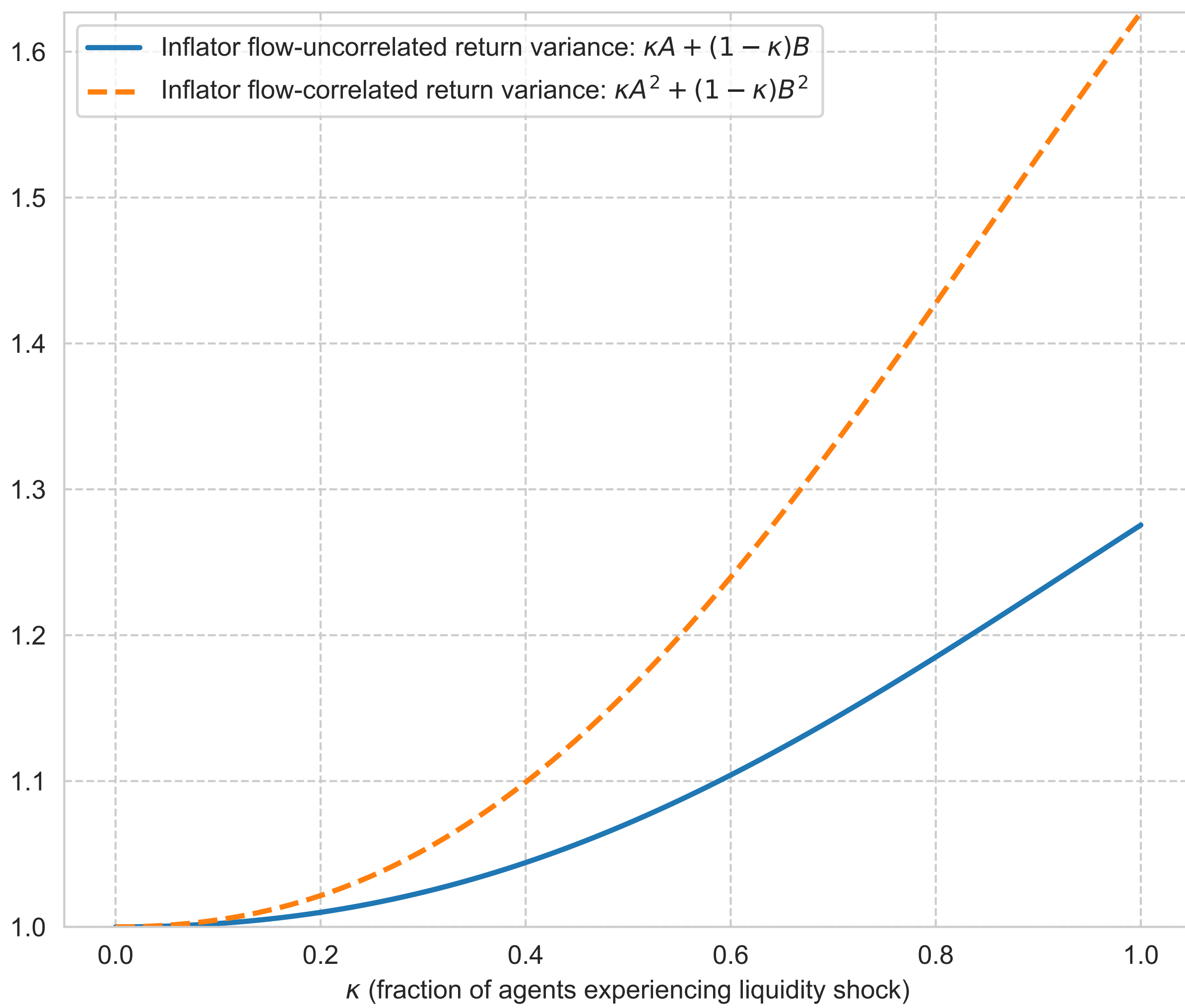

Figure 2: Variance inflators. This figure plots the variance inflators for the flow-correlated and the flow-uncorrelated returns. These inflators are used in the pricing of the variance swaps for $\alpha = 3$ and $\sigma^2 = 400$ percent squared annually. The flow-correlated part is assumed to be 20% of the total variance, and the flow-uncorrelated part 80% of it (i.e., $\sigma^2_{v_2} = 0.20\sigma^2$ and $\sigma^2_{v_3} = 0.80\sigma^2$). The variance of the liquidity shock is assumed to be nine: $\sigma^2_z = 9$.

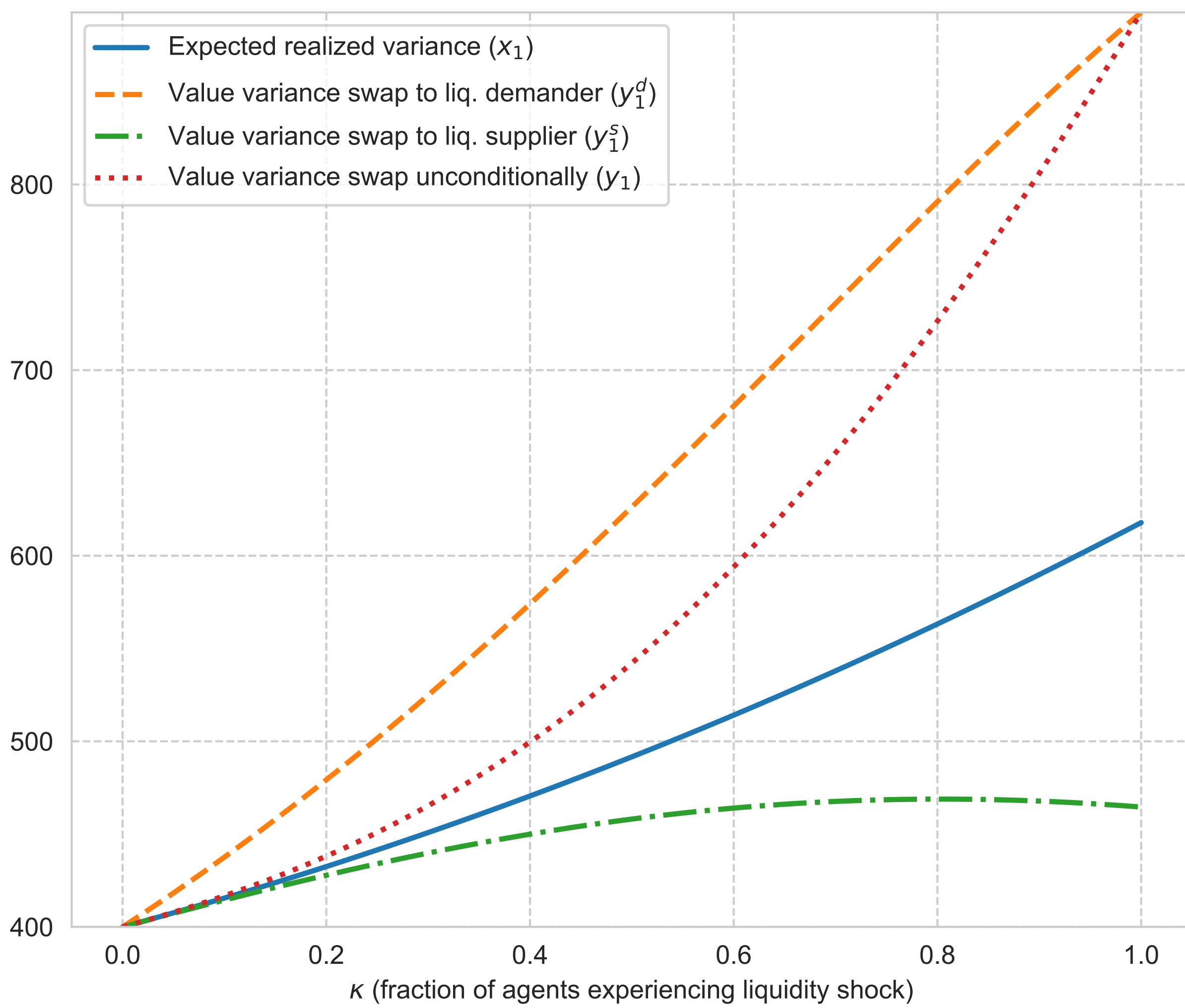

Figure 3: Pricing of the variance swap. This figure illustrates the pricing of the variance swaps for $\alpha = 3$ and $\sigma^2 = 400$ percent squared annually. The flow-correlated part is assumed to be 20% of the total variance, and the flow-uncorrelated part 80% of it (i.e., $\sigma_{v_2}^2 = 0.2\sigma^2$ and $\sigma_{v_3}^2 = 0.8\sigma^2$). The variance of the liquidity shock is assumed to be nine: $\sigma_z^2 = 9$. The graph plots the valuation of a variance swap by the two types. It also plots the pricing of the variance swap as the weighted sum of the type-specific valuations. The graph further plots expected realized variance.

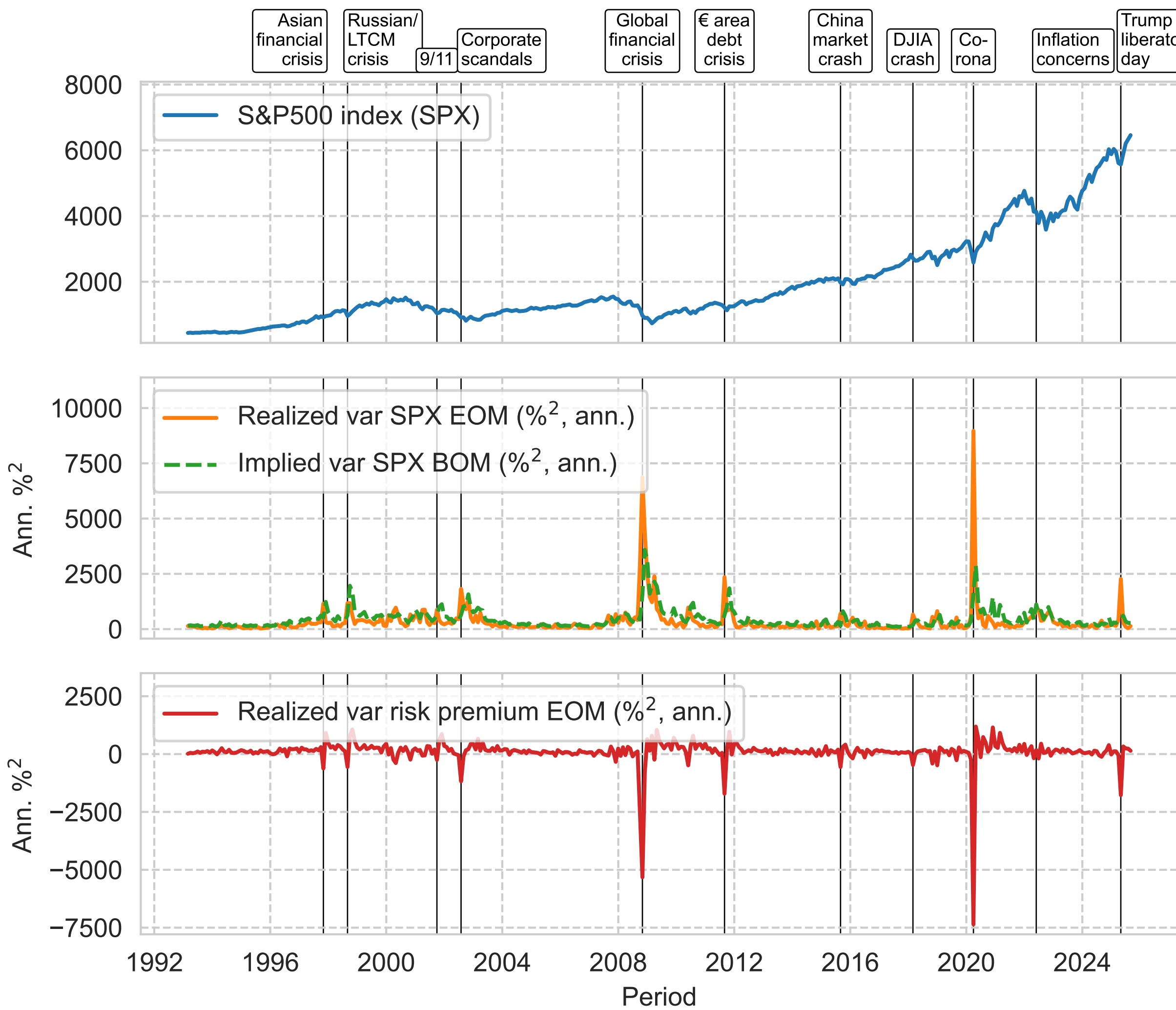

Figure 4: Time series S&P500 index returns and variances. This figure plots the S&P 500 index, the end-of-month realized variance of daily changes in this index, the beginning-of-month implied variance for these changes, and the wedge between the two, the realized variance risk premium.

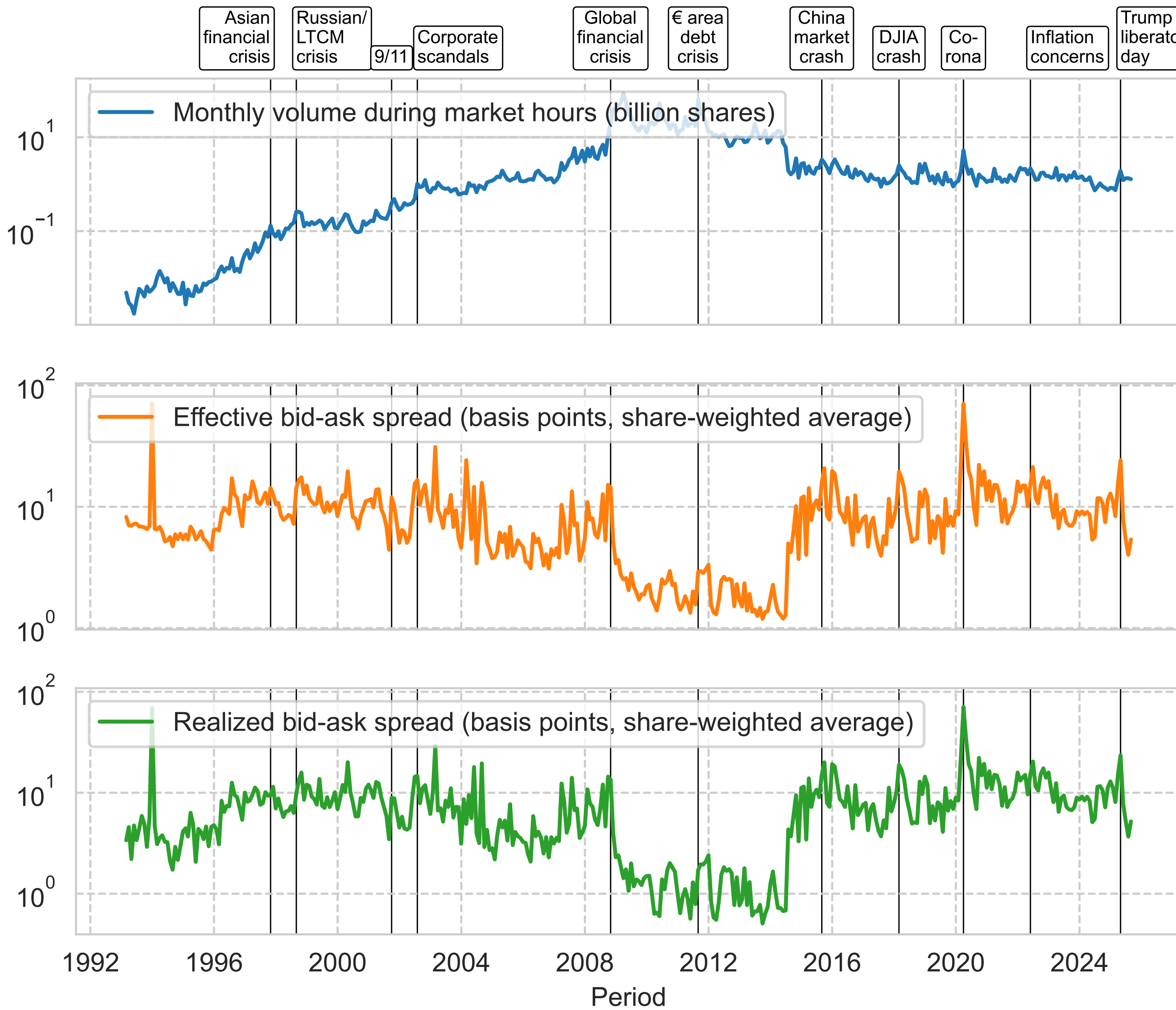

Figure 5: Time series SPY trading. This figure plots the following trade statistics for SPY: Volume, effective spread, and realized spread.

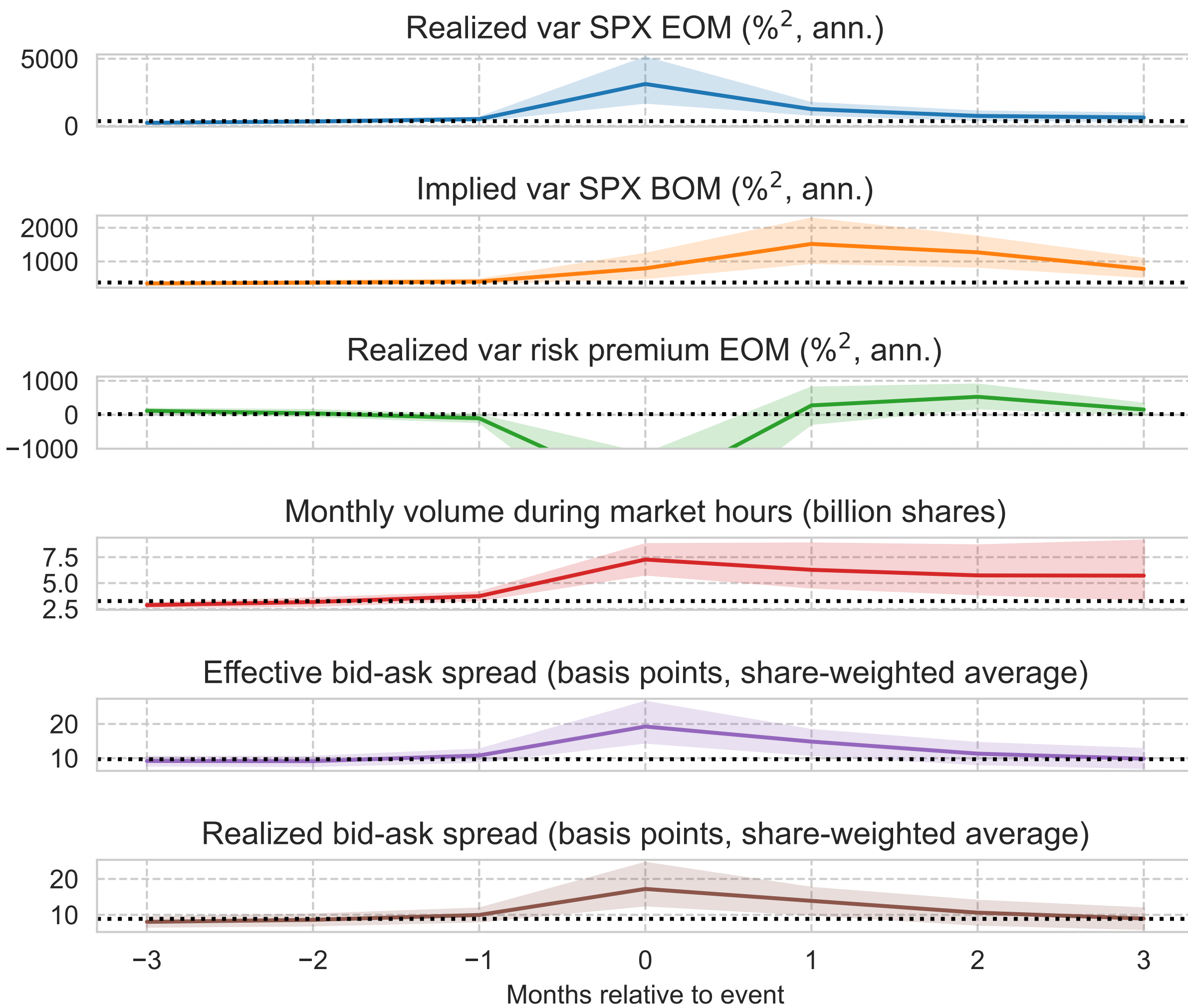

Figure 6: Crisis patterns. This figure plots how various time series evolve around crisis months. These crisis periods correspond with realized variance spikes (see vertical lines in Figure 4). All series are scaled by the pre-event level, which is set to 100. All plots show a 95% confidence interval.