Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets

1 Introduction

In this paper, we present new evidence connecting financial intermediary constraints to asset prices. We propose to measure the tightness of intermediary constraints based on how the intermediaries manage their aggregate tail risk exposures. Using data on the trading activities between public investors and financial intermediaries in the market of deep out-of-the-money put options on the S&P 500 index (abbreviated as DOTM SPX puts), we exploit the price-quantity relations to identify periods when shocks to intermediary constraints are likely to be the main driver of the variations in the net amount of trading between public investors and financial intermediaries. This then enables us to infer tightness of intermediary constraints from the option trading quantities. We show that a tightening of intermediary constraint according to our measure is associated with increasing option expensiveness, rising risk premia for a wide range of financial assets, deterioration in funding liquidity, as well as deleveraging by broker-dealers.

To construct the constraint measure, we start by computing the net amount of DOTM SPX puts that public investors in aggregate purchase each month (henceforth referred to as $PNBO$), which also reflects the net amount of the same options that broker-dealers and market-makers sell in that month. While it is well known that financial intermediaries are net sellers of these types of options during normal times, we find that $PNBO$ varies significantly over time and tends to fall/turn negative during times of market distress.

Periods when $PNBO$ is low could be periods of weak supply by intermediaries (due to tight constraints) or weak demand by public investors. One needs to separate these two effects in order to link $PNBO$ to intermediary constraints. We propose to exploit the relation between the quantities of trading, as measured by $PNBO$, and prices (expensiveness of SPX options), as measured by the variance premium. Positive comovements in prices and quantities are consistent with the presence of demand shocks, while negative comovements are consistent with the presence of supply shocks.[1] We summarize the daily

[1] We assume public investors' demand curve is downward sloping, and financial intermediaries' supply curve is upward sloping. These assumptions apply when both groups are (effectively) risk averse and cannot fully unload the inventory risks through hedging. Notice that the presence of one type of shocks does not rule out the presence of the other.

price-quantity relations each month. Those months with negative price-quantity relations on average are likely to be the periods when supply shocks are the main driver of the quantity of trading. Then, we take low $PNBO$ in a month with negative price-quantity relation as indicative of tight intermediary constraints.

In monthly data from January 1991 to December 2012, $PNBO$ is significantly negatively related to option expensiveness, and this negative relation becomes stronger when jump risk in the market is higher. In daily data, the correlation between $PNBO$ and our measure of option expensiveness is negative in 159 out of 264 months. These results highlight the significant role that supply shocks play in the market of DOTM SPX puts.

In those periods with a negative price-quantity relation, $PNBO$ significantly predicts future market excess returns. A one-standard deviation decrease in $PNBO$ (normalized $PNBO$) in a month with negative price-quantity relation is on average associated with a 4.3% (3%) increase in the subsequent 3-month log market excess return. The R^2 of the regression is 24.7% (12.3%). The predictive power of $PNBO$ is even stronger in the months when market jump risk is above the median level (in addition to the negative price-quantity relation), but becomes much weaker in the months when the price-quantity relations are positive. Besides equity, a lower $PNBO$ also predicts higher future excess returns for high-yield corporate bonds, an aggregate hedge fund portfolio, a carry trade portfolio, and a commodity index, and it predicts lower future excess returns on long-term Treasuries and (pay-fix) SPX variance swaps.

The predictability results survive an extensive list of robustness checks. They include different statistical methods for determining the significance of the predictive power, exclusion of the 2008-09 financial crisis and extreme observations of $PNBO$, different ways to define option moneyness, and an alternative quantity measure based on end-of-period open interest instead of trading volume, among others. In addition, we consider an alternative method for identifying periods of weak supply based on Rigobon (2003) (see also Sentana and Fiorentini (2001)). Using the reduced-form econometric assumptions of this method, we extract supply shocks to intermediaries and confirm the ability of the inferred supply shocks to predict future stock returns.

The return predictability results are consistent with the intermediary asset pricing theories, where a reduction in the risk-sharing capacity of the financial intermediaries causes the aggregate risk premium in the economy to rise. An alternative explanation of the predictability results is that $PNBO$ is merely a proxy for standard macro/financial factors that simultaneously drive the aggregate risk premium and intermediary constraints. If this alternative explanation is true, then the inclusion of proper risk factors into the predictability regression should drive away the predictive power of our constraint measure. We find that the predictive power of our measure is unaffected by the inclusion of a long list of return predictors in the literature, including various price ratios, consumption-wealth ratio, variance risk premium, default spread, term spread, and several tail risk measures. While these results do not lead to the rejection of the alternative explanation (there can always be omitted risk factors), they are at least consistent with intermediary constraints having a unique effect on the aggregate risk premium.

Our intermediary constraint measure is significantly related to the funding condition measures of Fontaine and Garcia (2012) (extracted from the Treasury market) and Adrian, Etula, and Muir (2014) (based on the growth rate of broker-dealer leverage). At the same time, we also find that our constraint measure provides unique information about the aggregate risk premium not contained in the other funding liquidity measures.

Our results suggest that when financial intermediaries switch from sellers of DOTM SPX puts to buyers (e.g., in the months following the Lehman Brothers bankruptcy in 2008), it is likely that the tightening of constraints are forcing the intermediaries to aggressively hedge their tail risk exposures, rather than the intermediaries accommodating an increase in public investors' demand to sell crash insurance. Examples of shocks to intermediary constraints include stricter regulatory requirements on banks' tail risk exposures (e.g., due to the Dodd-Frank Act or Basel III), or losses incurred by the intermediaries. To further examine the risk sharing mechanism, we try to identify who among the public investors (retail or institutional) are the "liquidity providers" during times of distress: reducing the net amount of crash insurance acquired from financial intermediaries or even providing insurance to the latter group. We answer this question by comparing public investors' demand in the markets of SPX vs. SPY options. SPY options are options on the

SPDR S&P 500 ETF Trust, which has a significantly higher percentage of retail customers than SPX options. Our results suggest that institutional public investors are the liquidity providers during periods of distress.

Our paper builds on and extends the work of Garleanu, Pedersen, and Poteshman (2009) (henceforth GPP) to incorporate the impact of supply shocks into the options market. In a partial equilibrium setting, GPP demonstrate how exogenous public demand shocks affect option prices when risk-averse dealers have to bear the inventory risks. In their model, the dealers' intermediation capacity is fixed, and the model implies a positive relation between the public demand for options and the option premium. Unlike GPP, we consider shocks to the intermediary constraint and the endogenous relations among public demand for options, option pricing, and aggregate market risk premium.[2] In the empirical analysis, we try to separate the effects of public demand shocks and shocks to intermediary constraints, and show that the latter is linked to the time-varying risk premia for a wide range of financial assets. Our empirical strategy based on the price-quantity dynamics is motivated by Cohen, Diether, and Malloy (2007), who use a similar strategy to identify demand and supply shocks in the equity shorting market.

The recent financial crisis has highlighted the importance of understanding the potential impact of intermediary constraints on the financial markets and the real economy. Following the seminal contributions by Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Bernanke, Gertler, and Gilchrist (1999), recent theoretical developments include Gromb and Vayanos (2002), Brunnermeier and Pedersen (2009), Geanakoplos (2009), He and Krishnamurthy (2013), Adrian and Boyarchenko (2012), Brunnermeier and Sannikov (2014), among others.

In contrast to the fast growing body of theoretical work, there is relatively little empirical work on measuring intermediary constraints and studying their aggregate effects on asset prices. The notable exceptions include Adrian, Moench, and Shin (2010) and Adrian, Etula, and Muir (2014), who show that changes in aggregate broker-dealer leverage is linked to the time series and cross section of asset returns. Our paper demonstrates a

[2] We present a general equilibrium model in the Online Appendix, which captures time-varying intermediary constraints in reduced form and is quite tractable.

new venue (the crash insurance market) to capture intermediary constraint variations and study their effects on asset prices. Moreover, compared to intermediary leverage changes, our measure has the advantage of being forward-looking and available at higher (daily and monthly instead of quarterly) frequency.

The ability of option volume to predict returns has been examined in other contexts. Pan and Poteshman (2006) show that option volume predicts near future individual stock returns (up to 2 weeks). They find the source of this predictability to be the nonpublic information possessed by option traders. Our evidence of return predictability applies to the market index and to longer horizons (up to 4 months), and we argue that the source of this predictability is time-varying intermediary constraints.

Finally, several studies have examined the role that derivatives markets play in the aggregate economy. Buraschi and Jiltsov (2006) study option pricing and trading volume when investors have incomplete and heterogeneous information. Bates (2008) shows how options can be used to complete the markets in the presence of crash risk. Longstaff and Wang (2012) show that the credit market plays an important role in facilitating risk sharing among heterogeneous investors. Chen, Joslin, and Tran (2012) show that the market risk premium is highly sensitive to the amount of sharing of tail risks in equilibrium.

2 Research Design

Our goal is to measure how constrained financial intermediaries are through the ways they manage their exposures to aggregate tail risks. The market of DOTM SPX put options are well-suited for this purpose. First, this market is large in terms of the economic exposures it provides for aggregate tail risks.[3] Second, compared to other over-the-counter derivatives that also provide exposures to aggregate tail risks, the exchange-traded SPX options have the advantages of better liquidity and almost no counterparty risk (other than exchange failure). Third, the Options Clearing Corporation (OCC) classifies exchange

[3]For example, based on the data in December 2012, Johnson, Liang, and Liu (2016) estimate that the change in value of index options outstanding is on the order of trillions of dollars following a severe market crash, the majority of which contributed by out-of-money SPX puts.

option transactions by investor types, which allows us to determine the net exposures of the financial intermediaries and is essential for constructing our measure.

Specifically, the OCC classifies each option transaction into one of three categories based on who initiates the trade. They include public investors, firm investors, and market-makers. Transactions initiated by public investors include those initiated by retail investors and those by institutional investors such as hedge funds. Trades initiated by firm investors are those that securities broker-dealers (who are not designated market-makers) make for their own accounts or for another broker-dealer. Since we focus on financial intermediaries as a whole, it is natural to merge firm investors and market-makers as one group and observe how they trade against public investors.

We classify DOTM puts as those with strike-to-price ratio $K/S \leq 0.85$. For robustness, we also consider different strike-to-price cutoffs, as well as cutoffs that adjust for option maturity and the volatility of the S&P 500 index (which is similar to cutoffs based on option delta). Another feature of option transaction is that an order can either be an open order (to open new positions) or a close order (to close existing positions). We will focus on open orders, because they are less likely to be mechanically influenced by existing positions (see Pan and Poteshman, 2006).

We construct a measure of the public net buying-to-open volume for DOTM SPX puts (abbreviated as $PNBO$). In period t (e.g., a day or a month), $PNBO_t$ is defined as

$$PBNO_t \equiv \text{public total open-buy volume}_t - \text{public total open-sell volume}_t. \quad (1)$$

$PNBO$ represents the amount of new DOTM SPX puts bought (sold if negative) by public investors in a period. Due to the growth in size of the options market, there could be a time trend in the level or volatility of $PNBO$. Thus, we also consider normalizing $PNBO$ by the average monthly volume of all SPX options traded by public investors over the past three months,[4]

$$PNBON_t \equiv \frac{PNBO_t}{\text{Average monthly public SPX volume over past 3 months}}. \quad (2)$$

While $PNBON$ helps address the potential issue with growth in the size of market, $PNBO$ has the advantage in that it better captures the actual magnitude of the tail risk exposures being transferred between public investors and intermediaries, which matters for measuring the degree of intermediary constraints. Considering this tradeoff, we conduct all of our main analyses using both $PNBO$ and $PNBON$.

It is well documented (see e.g., Bollen and Whaley, 2004) that, during normal times, public investors are net buyers of index puts while financial intermediaries are net sellers. All else equal, when financial intermediaries become more constrained, their willingness to supply crash insurance to the market will be reduced. It is thus tempting to infer how constrained financial intermediaries are based on the net amount of crash insurance they sell to public investors each period, as captured by $PNBO$. However, besides weak supply from constrained intermediaries, weak public demand can also cause the equilibrium amount of crash insurance traded between public investors and intermediaries to be low. The challenge is to separate the effects of supply from demand.

We address this problem in two ways, (1) by exploiting the price-quantity relation to identify periods of "supply environments," when variations in $PNBO$ are likely to be mainly driven by shocks to intermediary constraints, and (2) by the method of Rigobon (2003), which achieves identification by exploiting the heteroskedasticity of demand and supply shocks.

Before presenting the identification methodology, we briefly explain our "price" measure, i.e., the expensiveness of SPX options. One would ideally like to calculate the difference between the market price of an option and its hypothetical price without any market frictions. The latter is not observable and can only be approximated by adopting a specific pricing model. For simplicity and robustness, we use the variance premium (VP) in

[4]For robustness check, we also define $PNBON$ using past 12-month average public trading volume in the denominator, which generates similar results.

Bekaert and Hoerova (2014) as a proxy for overall expensiveness of SPX options, which is the difference between VIX^2 and the expected physical variance of the return of the S&P 500 index.[5]

2.1 Identification through the price-quantity relation

Our first empirical strategy is motivated by Cohen, Diether, and Malloy (2007) (CDM), who identify shifts in demand vs. supply in the securities shorting market by examining the relation between the changes in the loan fee (price) and the changes in the percentage of outstanding shares on loan (quantity). In their setting, a simultaneous increase (decrease) in the price and quantity indicates *at least* an increase (decrease) in shorting demand, whereas an increase (decrease) in price coupled with a decrease (increase) in quantity indicates *at least* a decrease (increase) in shorting supply.

The same logic applies to the options market. The demand pressure theory of GPP predicts that a positive exogenous shock to the public demand for DOTM SPX puts forces risk-averse dealers to bear more inventory risks. As a result, the dealers will raise the price of the option (a move along the upward-sloping supply curve). Thus, demand shocks generate a positive relation between changes in prices and quantities. Alternatively, if there are intermediation shocks that tighten the constraints facing financial intermediaries (e.g., due to loss of capital or higher capital requirements), they will become less willing to provide crash insurance to public investors. Then, the premium for the DOTM SPX puts rises while the equilibrium quantity of such options traded falls (a move along the downward-sloping demand curve).

Different from CDM, we would like to identify periods of weak and strong supply (level), instead of negative or positive supply shocks (changes). For this reason, we cannot directly apply their identification strategy. However, the ability to identify supply and demand shocks is still useful for our setting. Consider a month with low $PNBO$. If the

[5] The expected physical variance one month ahead (22 trading days) is computed using Model 8 in Bekaert and Hoerova (2014): $E_d\left[RV_{d+1}^{(22)}\right] = 3.730 + 0.108\frac{VIX_d^2}{12} + 0.199RV_d^{(-22)} + 0.33\frac{22}{5}RV_d^{(-5)} + 0.107 \cdot 22RV_d^{(-1)}$, where $RV_d^{(-j)}$ is the sum of daily realized variances from day $d-j+1$ to day d. The daily realized variance sums squared 5-minute intraday S&P500 returns and the squared close-to-open return.

price-quantity relations on a daily basis suggest mainly supply shocks in that month, then the low $PNBO$ is more likely to be driven by weak supply instead of weak demand.

Based on this idea, we run the following regression using daily data in each month t:

$$VP_{i(t)} = a_{VP,t} + b_{VP,t}\ PNBO_{i(t)} + d_{VP,t}\ J_{i(t)} + \epsilon^{v}_{i(t)}, \tag{3}$$

where $i(t)$ denotes day i in month t. The presence of jumps in the underlying stock index can affect VP even when markets are frictionless. Thus, when examining the relation between VP and $PBNO$, we control for the level of jump risk J in the S&P 500 index based on the measure of Andersen, Bollerslev, and Diebold (2007).

A negative coefficient $b_{VP,t} < 0$ in month t suggests that supply shocks are the dominant driver of price-quantity relations in that month, and we expect $PNBO$ to be informative about the variation in intermediary constraints during such times. The fact that $b_{VP,t} < 0$ does not identify any particular supply shock, nor does it rule out the presence of demand shocks in the same month. It does indicate that supply shocks are likely to be more significant relative to demand shocks. Similarly, $b_{VP,t} > 0$ does not rule out the presence of supply shocks in a month, but demand shocks are likely to be more significant. During such periods, we do not expect $PNBO$ to be informative about supply conditions.

Furthermore, we expect high jump risk to amplify the effect of shocks to intermediary constraints on the equilibrium quantity of options traded. This is because a main reason that intermediary constraint matters for their supply of DOTM index puts is the difficulty to hedge the market jump risk embedded in their inventory positions. If public demand does not become more volatile during such times, variations in $PNBO$ will be more informative about shocks to intermediary constraints when jump risk is high. Whether this assumption is valid or not is an empirical question.

In summary, in a month when the price-quantity relation is on average negative ($b_{VP,t} < 0$), we expect small (or negative) value for $PNBO_t$ (cumulative net-buying by public investors for the month) to indicate tight intermediary constraints.

2.2 Identification through heteroskedasticity

Besides the identification of "supply environments" based on the price-quantity relation, an alternative method to identify supply shocks is through econometric identification. One method suited for our study is identification through heteroskedasticity of Rigobon (2003).[6] Rigobon (2003) considers a standard linear supply-demand relationship between prices (p_t) and quantities (q_t):

$$p_t = b + \beta q_t + \epsilon_t, \quad \text{(demand equation)} \tag{4a}$$

$$q_t = a + \alpha p_t + \eta_t, \quad \text{(supply equation)} \tag{4b}$$

where the volatilities of the supply and demand shocks are σ_ϵ and σ_η, respectively. In general, the residuals will be correlated with the independent variables in each equation and the parameters will not be identified. Rigobon (2003) solves this identification problem by considering regime-dependent heteroscedasticity of (ϵ, η). Supposing that there are two regimes and the relative volatilities of the supply and demand shocks vary across the regimes, the supply and demand equations can be identified.

In parallel to our empirical strategy motivated by CDM, we also identify demand and supply shocks following the method of Rigobon (2003). There is suggestive evidence that in the market for DOTM SPX puts supply shocks are more volatile relative to demand shocks during the period of the U.S. financial crisis and the European sovereign debt crisis. We thus date the low and high-supply volatility regimes accordingly, and use the price and quantity measures discussed above to estimate the supply-demand system in (4a–4b).

The two identification methods presented in Section 2.1 and 2.2 have their respective advantages. On the one hand, the Rigobon method has the advantage of being based on a clean set of parametric assumptions, which helps with precise identification of the supply equation and supply shocks. However, these assumptions might be restrictive and potentially inconsistent with the data (e.g., the assumption about the linear relations between prices and quantities, the number of volatility regimes, and whether other

[6] We thank an anonymous referee for this suggestion.

parameters might change across these regimes, etc). On the other hand, the method based on price-quantity relation does not identify the supply environment or supply shocks as cleanly, but it imposes weaker assumptions on the demand and supply curve that likely makes the results more robust.

After constructing our measures of intermediary constraints, we investigate how these measures are linked to asset prices. According to the theory of financial intermediary constraints (see e.g., Gromb and Vayanos (2002) and He and Krishnamurthy (2013)), variations in the aggregate intermediary constraints not only affect option prices, but also drive the risk premia of other financial assets. This theory implies that low $PNBO_t$, when occurring in a period dominated by supply shocks, should imply high future expected excess returns on the market portfolio. That is, we expect $b_r^- < 0$ in the following predictive regression:

$$r_{t+j\to t+k} = a_r + b_r^- \ I_{\{b_{VP,t}<0\}} \, PNBO_t + b_r^+ \ I_{\{b_{VP,t}\geq 0\}} \, PNBO_t + c_r \ I_{\{b_{VP,t}<0\}} + \epsilon_{t+j\to t+k} \quad (5)$$

where r denotes log market excess return, and the notation $t + j \to t + k$ indicates the leading period from $t + j$ to $t + k$ ($k > j \geq 0$). Similarly, we expect $\hat{\eta}_t$ extracted from the supply-demand system in (4a–4b) to predict future market excess returns as well,

$$r_{t+j\to t+k} = a_s + b_s \hat{\eta}_t + \epsilon_{t+j\to t+k} \quad (6)$$

where we expect $b_s \leq 0$. Besides the market portfolio, the predictability should apply to other risky assets as well.

Finally, the empirical strategy and testable hypotheses above are mainly based on economic intuition. In the Online Appendix, we present a dynamic general equilibrium model featuring time-varying intermediary constraints. The model not only helps formalize the main intuition, but generates more rigorous predictions about how intermediary constraints affect the equilibrium price-quantity dynamics in the crash insurance market, the aggregate risk premium, and intermediary leverage. Moreover, we can use the calibrated model to examine the quantitative effects of intermediary constraints on asset prices.

3 Empirical Results

We now present the empirical evidence connecting the option trading activities to the constraints of the financial intermediaries and the risk premia in financial markets.

3.1 Data

Figure 1 plots the monthly time series of $PNBO$ and its normalized version $PNBON$. Consistent with the finding of Pan and Poteshman (2006) and GPP, the net public purchase of DOTM SPX puts was positive for the majority of the months prior to the financial crisis in 2008, suggesting that broker-dealers and market-makers were mainly supplying crash insurance to public investors. A few notable exceptions include the period around the Asian financial crisis (December 1997), Russian default and the financial crisis in Latin America (November 1998 to January 1999), the Iraq War (April 2003), and two months in 2005 (March and November 2005).[7]

However, starting in 2007, $PNBO$ became significantly more volatile.[8] It turned negative during the quant crisis in August 2007, when a host of quant-driven hedge funds experienced significant losses. It then rose significantly and peaked in October 2008, following the Lehman Brothers bankruptcy. As market conditions continued to deteriorate, $PNBO$ plunged rapidly and turned significantly negative in the following months. Following a series of government interventions, $PNBO$ bottomed in April 2009, rebounded briefly, and then dropped again in December 2009 when the Greek debt crisis escalated. During the period from November 2008 to December 2012, public investors on average sold 44,000 DOTM SPX puts to open new positions each month. In contrast, they bought on average 17,000 DOTM SPX puts each month in the period from 1991 to 2007.

One reason that the $PNBO$ series appears more volatile in the latter part of the sample is that the options market (e.g., in terms of total trading volume) has grown significantly

[7]The GM and Ford downgrades in May 2005 might be related to the negative $PNBO$ in 2005.

[8]One potential concern is that the volatility of $PNBO$ may be non-stationary. If we apply the structural break test of Andrews (1993) to the monthly realized volatility (computed from daily $PNBO$), we find evidence for a break in the time trend in volatility at $\tau =$ March 2007 ($F_{max} = 50.4 > F_{crit,1\%} = 15.56$). Subsequent to this break, the point estimate of the time trend is negative, indicating that volatilities declined (or at least didn't continue to increase) from this point.

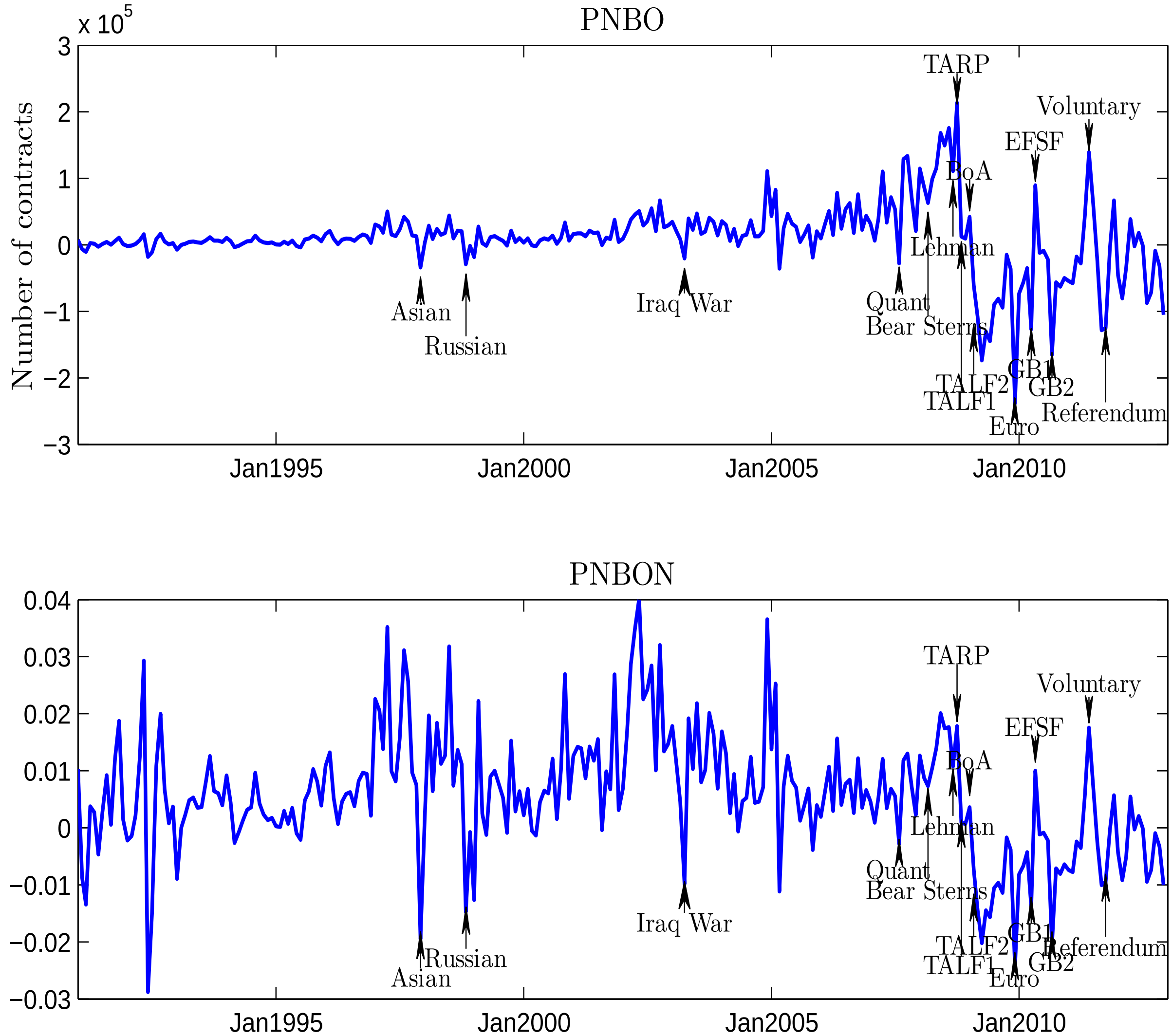

Figure 1: **Time Series of net public purchase for DOTM SPX puts.** $PNBO$ is the net amount of deep-out-of-the-money (DOTM)(with $K/S \leq 0.85$) SPX puts public investors buying-to-open each month. $PNBON$ is $PNBO$ normalized by average of previous 3-month total volume from public investors."Asian" (1997/10): period around the Asian financial crisis. "Russian" (1998/11): period around Russian default. "Iraq" (2003/04): start of the Iraq War. "Quant" (2007/08): the crisis of quant-strategy hedge funds. "Bear Sterns" (2008/03): acquisition of Bear Sterns by JPMorgan. "Lehman" (2008/09): Lehman bankruptcy. "TARP" (2008/10): establishment of TARP. "TALF1" (2008/11): creation of TALF. "BoA" (2009/01): Treasury, Fed, and FDIC assistance to Bank of America. "TALF2" (2009/02): increase of TALF to $1 trillion. "Euro" (2009/12): escalation of Greek debt crisis. "GB1" (2010/04): Greece seeks financial support from euro and IMF. "EFSF" (2010/05): establishment of EFSM and EFSF; 110 billion bailout package to Greece agreed. "GB2" (2010/09): a second Greek bailout installment. "Voluntary" (2011/06): Merkel agrees to voluntary Greece bondholder role. "Referendum" (2011/10): further escalation of Euro debt crisis with the call for a Greek referendum.

over time. As the bottom panel of Figure 1 shows, after normalizing $PNBO$ with the total SPX volume (see the definition in (2)), the $PNBON$ series no longer demonstrates visible trend in volatility.

Table 1 reports the summary statistics of the option volume and pricing variables and their correlation coefficients. From January 1991 to December 2012, the public net buying-to-open volume of DOTM SPX puts ($PNBO$) is close to 10,000 contracts per month on average (each contract has a notional size of 100 times the index). In comparison, the average total open interest for all DOTM SPX puts is around 0.9 million contracts during the period from January 1996 to December 2012, which highlights the significant difference between $PNBO$ and open interest. The option volume measures have relatively modest autocorrelations at monthly frequency (0.61 for $PNBO$ and 0.48 for $PNBON$) compared to standard return predictors such as dividend yield and term spread. The correlation matrix in Panel B shows that the various quantity measures are negatively related to variance premium (VP). In addition, both $PNBO$ and $PNBON$ are negatively correlated with the unemployment rate (see Table IA1 in the Internet Appendix).

Figure 2 provides information about the trading volume of SPX options at different moneyness. Over our entire sample, put options account for 63% of the total trading volume of SPX options. Among put options, out-of-the-money puts account for over 75% of the total trading volume; in particular, DOTM puts (with $K/S < 0.85$) account for 23% of the total volume. These statistics demonstrate the importance of the market for DOTM SPX puts.

While financial intermediaries can partially hedge the risks of their option inventories through dynamic hedging, the hedge is imperfect and costly. This is especially true for DOTM SPX puts, because they are highly sensitive to jump risk that are difficult to hedge. To demonstrate this point, we regress put option returns on the returns of the corresponding hedging portfolios at both weekly and daily horizons. We consider both delta hedging (using the S&P 500 index) and delta-gamma hedging.[9] The R^2 of these regressions demonstrate how effective the hedging methods are.

[9] We restrict the options to be between 15 and 90 days to maturity to ensure liquidity. For delta-gamma hedging, we use at-the-money puts expiring in the following month in addition to the S&P500 index.

Table 1: **Summary Statistics and Correlation Coefficients**

This table reports the summary statistics for the SPX options volume from public investors and pricing variables in the empirical analysis. $PNBO$: public net open-buying volume of DOTM puts ($K/S \leq 0.85$). $PNBON$: $PNBO$ normalized by average monthly public SPX volume over past 12 months (in million contracts). $PNBO_{ND}$: public net open-buying volume of all SPX options excluding DOTM puts. $PNOI$: public net open interest for DOTM SPX puts (in million contracts). $PNOIN$: $PNOI$ normalized by the total public open interest of all options (long and short). J: monthly average of the daily physical jump risk measure by Andersen, Bollerslev, and Diebold (2007). VP: variance premium based on Bekaert and Hoerova (2014). AC(1) is the first order autocorrelation for monthly time series; pp-test is the p-value for the Phillips-Perron test for unit root. Sample period: 1991/01 – 2012/12.

Panel A: Summary Statistics

	mean	median	std	AC(1)	pp-test
$PNBO$ (10^3 contracts)	10.00	9.67	51.12	0.61	0.00
$PNBON$ (%)	0.56	0.53	1.07	0.48	0.00
$PNBO_{ND}$ (10^3 contracts)	138.08	117.10	113.64	0.76	0.01
$PNOI$ (10^3 contracts)	28.79	19.85	63.21	0.74	0.00
$PNOIN$ (%)	0.18	0.15	0.36	0.70	0.00
J (%)	12.14	10.81	6.17	0.62	0.00
VP	19.37	14.56	21.66	0.54	0.00

Panel B: Correlation

	$PNBO$	$PNBON$	$PNBO_{ND}$	$PNOI$	$PNOIN$	J
$PNBON$	0.71					
$PNBO_{ND}$	0.21	0.04				
$PNOI$	0.67	0.48	0.33			
$PNOIN$	0.58	0.59	0.23	0.90		
J	0.03	0.03	0.08	0.16	0.12	
VP	−0.22	−0.15	−0.06	−0.07	−0.09	0.54

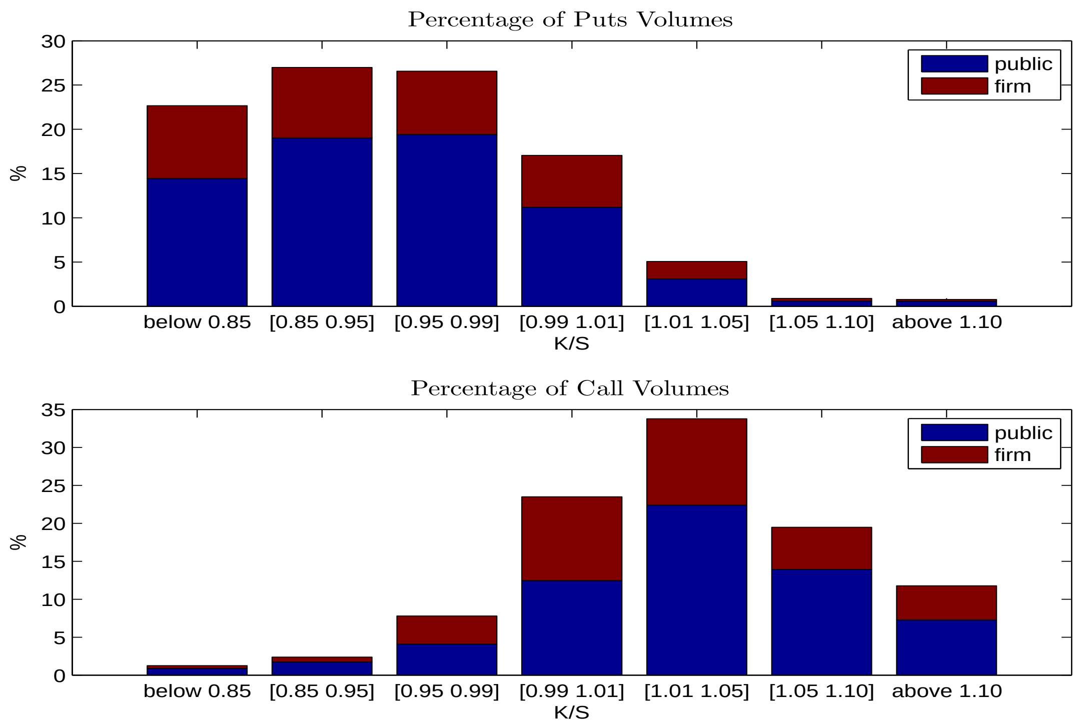

Figure 2: **Percentage of total put and call volumes at different moneyness.** This figure plots the total fraction of volume for calls and puts at different levels of moneyness (measure by strike price, K, divided by spot price S). The height of each bar indicates the fraction of the total market volume at that moneyness level, while the colors indicate the breakdown within the strike between public (blue) and private (red) orders.

As Table 2 shows, with daily (weekly) rebalancing, delta hedging can capture around 72% (76%) of the return variation of ATM SPX puts, but only 41% (34%) of the return variation of DOTM puts. With delta-gamma hedging, the R^2 for ATM puts can exceed 90%, but it is still below 60% for DOTM puts. These results imply that when holding non-zero inventories of DOTM SPX puts, financial intermediaries will be exposed to significant inventory risks even after dynamically hedging these positions. It is because of such inventory risks that financial intermediaries become more reluctant to supply crash insurance to the public investors when they are more constrained.

Table 2: **Explaining Options Returns with Hedging Portfolios**

This table shows the R^2 from regressing option returns on hedging portfolios returns. The dependent variables are the returns of put options with different moneyness. delta denotes the returns on the delta hedging portfolio for the corresponding put option. delt+gam denotes the returns on the delta-gamma hedging portfolio. Sample period: 1996 – 2012.

	$\frac{K}{S} < 0.85$		$0.85 < \frac{K}{S} < 0.95$		$0.95 < \frac{K}{S} < 0.99$		$0.99 < \frac{K}{S} < 1.01$	
	delta	delt+gam	delta	delt+gam	delta	delt+gam	delta	delt+gam
Weekly R^2	0.34	0.54	0.45	0.75	0.59	0.86	0.76	0.91
Daily R^2	0.41	0.59	0.46	0.74	0.56	0.82	0.72	0.87

3.2 Option volume and the expensiveness of SPX options

We start by investigating the link between $PNBO$ and the expensiveness of SPX options as proxied by the variance premium (VP) in Bekaert and Hoerova (2014). Before constructing the measure $b_{VP,t}$ in Equation (3) for the price-quantity relation based on daily data, we first examine the relation between $PNBO$ and VP at monthly frequency.

Table 3 reports the results. In both the cases of $PNBO$ and $PNBON$, the coefficient b_{VP} is negative and statistically significant, consistent with the hypothesis that shocks to intermediary constraints generate a negative relation between the equilibrium quantities of DOTM SPX puts that public investors purchase and the expensiveness of SPX options. The coefficient (-94.60) in the univariate regression suggests that a one standard deviation decrease in $PNBO$ is associated with an increase in VP of 4.82, a 25% increase relative to the average variance premium. Similarly, a one standard deviation decrease in $PNBON$ is associated with a 3.15 unit increase in VP.

After adding the interaction between $PNBO$ and the jump risk measure J into the regression, the coefficient c_{VP} of the interaction term is significantly negative, which implies that $PNBO$ and VP are more likely to be negatively related during times of high jump risk, and that their relation can turn positive when jump risk is sufficiently low. To understand the economic magnitude of the coefficient c_{VP}, we can compare how the marginal effect of $PNBO$ on VP, $b_{VP} + c_{VP}J$, changes for different levels of jump risk. Our estimates imply that a one standard deviation decrease in $PNBO$ is associated

Table 3: $PNBO$ **and SPX Option Expensiveness**

The dependent variable is VP. We use three different measures of the public net-buying volumes: $PNBO$, $PNBON$, and $PNBO_{ND}$ (public net buying volume of all SPX options excluding deep-out-of-the-money puts). Standard errors in parentheses are computed based on Newey and West (1987) with 6 lags. (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$VP_t = a_{VP} + b_{VP}\ PNBO_t + c_{VP}\ J_t \times PNBO_t + d_{VP}\ J_t + \epsilon_t^v$$

	$PNBO$			$PNBON$			$PNBO_{ND}$		
a_{VP}	20.32***	-2.83	-4.80	21.05***	-1.88	-5.92	20.96***	-1.13	-7.08
	(2.60)	(3.05)	(3.17)	(3.32)	(3.35)	(3.45)	(3.35)	(4.23)	(7.40)
b_{VP}	-94.60**	-101.57***	13.16	-2.94**	-3.24***	4.51	-11.48	-20.22	11.43
	(40.18)	(24.13)	(27.09)	(1.39)	(1.40)	(2.80)	(18.16)	(13.73)	(29.18)
c_{VP}			-6.83***			-0.57**			-2.33
			(1.81)			(0.24)			(2.85)
d_{VP}		1.91***	2.05***		1.90***	2.20***		1.92***	2.38***
		(0.31)	(0.32)		(0.35)	(0.34)		(0.38)	(0.75)
R^2	4.6	34.1	36.8	1.7	30.9	34.5	0.0	29.5	30.1

with an increase in VP of 1.41 when J is one standard deviation below its mean, while the increase in VP is 5.71 when J is one standard deviation above its mean. Similar calculations show that the marginal effect of $PNBON$ on VP is even more sensitive to J. The result is again consistent with the intermediary constraint theory, as the effects of shocks to intermediary constraints on the supply of DOTM SPX puts by financial intermediaries tend to strengthen when the aggregate tail risk is high.

In contrast, when we replace $PNBO$ with the public net buying volume for all other SPX options excluding DOTM puts ($PNBO_{ND}$), not only are the R^2 of the regressions smaller, but the regression coefficients b_{VP} and c_{VP} are no longer significantly different from zero. As we have demonstrated in Table 2, DOTM SPX puts are more difficult to hedge and hence expose intermediaries to higher inventory risks. Thus, the trading activities of DOTM SPX puts are likely to be more informative about the fluctuations in intermediary constraints compared to those for other options.

Garleanu, Pedersen, and Poteshman (2009) shows that exogenous public demand shocks can generate a positive relation between net public demand for index options and measures of option expensiveness. They find support for this prediction using data from October 1997 to December 2001. Bollen and Whaley (2004) also find evidence of the effects of public demand pressure on option pricing in daily data.

Our results above are not a rejection of the effect of demand shocks on option prices. In Section 4.5, we replicate the results of Table 2 in GPP and show that the different time periods is the main reason for the opposite signs of the price-quantity relation in the two papers. Conceptually, the demand pressure theory and the intermediary constraint theory share the common assumption of constrained intermediaries, and both can be at work in the data. For instance, our results in Table 3 show that the price-quantity relation is more likely to be negative (positive) when the jump risk in the market is high (low), indicating that the supply (demand) effects tend to become dominant under such conditions.

Next, we estimate the monthly price-quantity relation measure $b_{VP,t}$ from regression (3) using daily data. The fact that demand effects and supply effects are both present in the data is again evident. Out of 264 months, the coefficient $b_{VP,t}$ is negative in 159 (significant at 5% level in 44 of them), and positive in 105 (significant at 5% level in 24 of them). These statistics suggest that, according to the price-quantity relation, shocks to intermediary constraints are present in a significant part of our sample period. The months that have significantly negative price-quantity relations include periods in the Asian financial crisis, Russian default, the 2008 financial crisis, and several episodes during the European debt crisis.[10]

3.3 Option volume and risk premia

We now examine the predictions from Section 2 linking $PNBO$ and risk premia in the financial markets.

For initial exploration, we run the basic univariate return-forecasting regression using $PNBO$ and $PNBON$. Table 4 shows that $PNBO$ has strong predictive power for future market excess returns up to 4 months ahead. The coefficient estimate b_r for predicting

[10]Table A1 in the Appendix provides more details, as well as a comparison between our strategy and a direct application of the CDM method to identify supply environments.

Table 4: **Return Forecasts with $PNBO$**

This table reports the results of the return forecasting regressions using $PNBO$ and $PNBON$. $r_{t+j\to t+k}$ represents log market excess return from month $t+j$ to $t+k$ ($k > j \geq 0$). Standard errors in parentheses are computed based on Hodrick (1992). (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$r_{t+j\to t+k} = a_r + b_r\ PNBO_t + \epsilon_{t+j\to t+k}$$

Horizon	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2
	$PNBO$			$PNBON$		
$r_{t\to t+1}$	-24.26***	(7.04)	7.7	-0.92***	(0.32)	4.8
$r_{t+1\to t+2}$	-19.03***	(6.79)	4.8	-0.65**	(0.28)	2.5
$r_{t+2\to t+3}$	-23.86***	(7.64)	7.5	-0.75***	(0.29)	3.2
$r_{t+3\to t+4}$	-18.05**	(7.65)	4.3	-0.62**	(0.29)	2.2
$r_{t\to t+3}$	-67.16***	(18.96)	18.0	-2.32***	(0.75)	9.4

one-month ahead market excess returns is -24.26 (t-stat of -3.45),[11] with R^2 of 7.7%. For 4-month ahead returns ($r_{t+3\to t+4}$), the coefficient estimate is -18.05 and statistically significant (t-stat of -2.36), and R^2 drops to 4.3%. From 5 months out, the predictive coefficient is no longer statistically significant. When we aggregate the effect for the cumulative market excess returns in the next 3 months, the coefficient b_r is -67.16 (t-stat of -3.54) and R^2 is 18.0%. The economic significance that this coefficient estimate implies is striking. A one-standard deviation decrease in $PNBO$ is associated with a 3.4% (non-annualized) increase in the future 3-month market excess return.

Figure 1 indicates that non-stationarity might be a potential concern for $PNBO$. The autocorrelation of $PNBO$ is only 0.61 and a Phillips-Perron test strongly rejects the null of a unit root (see Table 1). However, non-stationarity may arise elsewhere, e.g., through the 2nd moment. For this reason, we also use the normalized $PNBO$ to predict market excess returns. Table 4 shows that, like $PNBO$, $PNBON$ also predicts future market returns negatively. The coefficient estimate b_r remains statistically significant

[11] All the standard errors for the return-forecasting regressions are based on Hodrick (1992). We provide additional results on statistical inference in Table IA2 in the Internet Appendix, including Newey and West (1987) standard errors with long lags, bootstrapped confidence intervals, and the test statistic of Muller (2014). See Ang and Bekaert (2007) for further discussion on long-horizon statistical inference.

up to 4 months ahead, but with lower R^2 than $PNBO$ at all horizons. The difference in R^2 between $PNBON$ and $PNBO$ shows that we should interpret the high R^2 for $PNBO$ with caution, which could partially be due to its volatility trend. As for economic significance, a one-standard deviation decrease in $PNBON$ is associated with a 2.5% increase in the future 3-month market excess return.

While $PNBO$ shows predictive power for market risk premia in the full sample, theories of intermediary constraints imply that the predictive power should be concentrated in periods when variations in $PNBO$ are mainly driven by changes in supply conditions. In Section 2, we have proposed to identify supply environments based on the negative price-quantity relation ($b_{VP,t} < 0$), and we expect high level of jump risks (J_t) to also help with identifying such environments under certain assumptions. Next, we examine how the predictive power of $PNBO$ changes in the sub-samples identified by $b_{VP,t}$ and J_t.

Table 5 shows that the predictive power of both $PNBO$ and $PNBON$ are indeed stronger (in terms of both economic and statistical significance) during the periods of negative price-quantity relation and during periods of high jump risks.[12] For example, when $b_{VP,t} < 0$, a one-standard deviation decrease in $PNBO$ ($PNBON$) is associated with a 4.1% (3.0%) increase in the future 3-month market excess return. The coefficient for $PNBO$ becomes considerably smaller in the sub-sample when $b_{VP,t}$ is positive, and it becomes insignificant for $PNBON$.

We then further split the full sample into 4 sub-sample periods based on the two criteria ($b_{VP,t} < (\geq) 0$, $J_t < (\geq) \bar{J}$).[13] Table 5 shows that for both $PNBO$ and $PNBON$, the predictive power is the strongest when $b_{VP,t} < 0$ and the level of jump risk is high. If $b_{VP} \geq 0$ and jump risk is low, then the coefficient b_r becomes positive and insignificant for both measures, and the R^2 drops to near zero.

In summary, the sub-sample results suggest that our strategy based on the price-quantity relation does a good job identifying those periods when $PNBO$ are connected to

[12]We set $\bar{J}$ to the median for J_t in the full sample. This potentially introduces future information into the return-forecasting regression. Our results are robust to changing $\bar{J}$ to only using past information.

[13]While using the level of jump risk to split the sample is motivated by the difficulty to hedge tail risk, the correlation between the two dummy variables of whether J_t and VP_t are above their respective sample medians is 0.97 (the correlation between J_t and VP_t is 0.54), meaning we will obtain essentially the same results if we split the samples based on VP_t.

Table 5: **Return Forecasts with $PNBO$: Sub-sample Results**

This table reports the sub-sample results of the 3-month return forecasting regressions. Standard errors in parentheses are computed based on Hodrick (1992). (***,** ,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$r_{t\to t+3} = a_r + b_r \ PNBO_t + \epsilon_{t+j\to t+k}$$

Sub-sample	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2	obs
	$PNBO$			$PNBON$			
$b_{VP,t} < 0$	-84.75***	(26.16)	24.7	-2.67***	(0.85)	12.3	159
$b_{VP,t} \geq 0$	-42.46**	(19.65)	9.2	-1.58	(1.01)	4.3	105
$J_t \geq \bar{J}$	-87.49***	(26.26)	26.8	-3.08***	(0.98)	13.0	132
$J_t < \bar{J}$	-33.80*	(17.93)	6.0	-1.47*	(0.81)	5.6	132
$b_{VP,t} < 0, J_t \geq \bar{J}$	-114.11***	(42.65)	32.4	-3.55***	(1.19)	14.2	80
$b_{VP,t} < 0, J_t < \bar{J}$	-53.19***	(18.61)	16.3	-1.95**	(0.91)	12.1	79
$b_{VP,t} \geq 0, J_t \geq \bar{J}$	-60.98***	(22.39)	20.8	-2.36**	(1.16)	10.5	52
$b_{VP,t} \geq 0, J_t < \bar{J}$	17.53	(37.12)	1.3	0.50	(1.74)	0.4	53

variations in intermediary constraints and in turn the conditional market risk premia. For the remainder of the paper, we use the regression specification (5), which summarizes the sub-sample results succinctly.

To investigate whether the predictability results above are useful in forming real-time forecasts, we follow Welch and Goyal (2008) and compute the out-of-sample R^2 for $PNBO$ and $PNBON$ based on various sample-split dates, starting in January 1996 (implying a minimum estimation period of 5 years) and ending in December 2007 (with a minimum evaluation period of 5 years). We consider the wide range of sample-split dates because recent studies suggest that sample splits themselves can be data-mined (Hansen and Timmermann (2012)). In forming the return forecasts, we first estimate the predictability regression (5) during the estimation period (from date 1 to t), and then use the estimated coefficients to forecast the 3-month future market excess return for $t+1$. After obtaining all the return forecasts, we then compute the mean squared forecast errors for the predictability model (MSE_A) and the historical mean model (MSE_N) in various

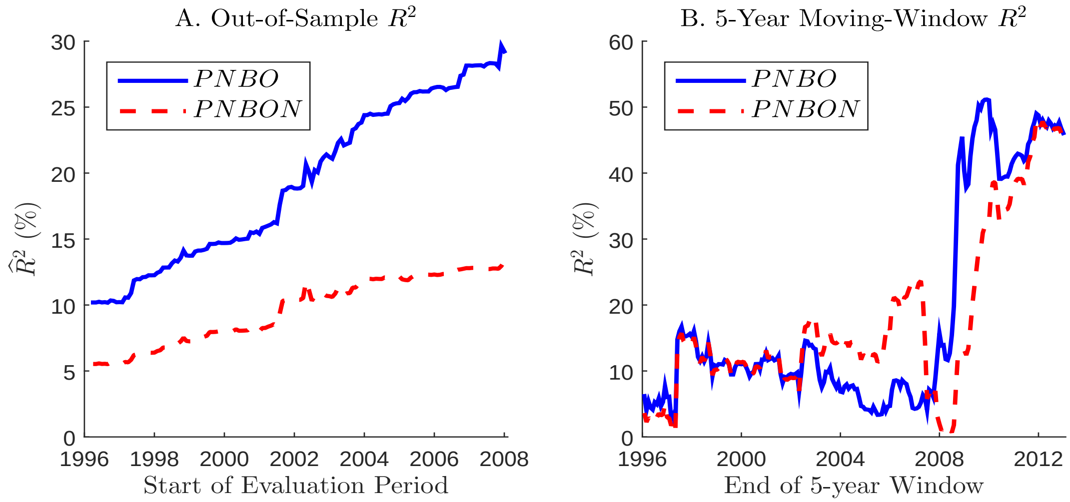

Figure 3: **Out-of-sample R^2 and R^2 from 5-year moving-window regressions.** This figures plots the out-of-sample R^2 and R^2 of 5-year moving windows from forecasting 3-month market excess returns based on the specification of (5). Panel A plots the out-of-sample R^2 as a function of the sample split date. Panel B plots the in-sample R^2 as a function of the end of 5-year moving windows.

evaluation periods that begin at the sample-split dates and end at the end of the full sample. The out-of-sample R^2 is given by

$$\widehat{R}^2 = 1 - \frac{MSE_A}{MSE_N}.$$

Panel A of Figure 3 shows the results. $PNBO$ achieves an out-of-sample R^2 above 10% for all the sample splits and remains above 20% from 2003 onward. $PNBON$ has an an out-of-sample R^2 above 5% for all the sample splits and remains above 10% in the later period. All of the out-of-sample R^2 are significant at the 5% level (1% level since 2000) based on the MSE-F statistic by McCracken (2007).

Panel B of Figure 3 plots the in-sample R^2 from the predictive regressions of $PNBO$ and $PNBON$ using 5-year moving windows. The two R^2s vary significantly over time. They are low at the beginning of the sample. Both R^2 rise to near 18% in the period around the Asian financial crisis and Russian default in 1997-98. During the 2008-9 financial crisis period, the R^2 rise above 40%. These high R^2 for the return-forecasting regressions would translate into striking Sharpe ratios for investment strategies that try to exploit such predictability. For example, Cochrane (1999) shows that the best unconditional Sharpe ratio s^* for a market timing strategy is related to the predictability regression R^2 by

$$s^* = \frac{\sqrt{s_0^2 + R^2}}{\sqrt{1 - R^2}},$$

where s_0 is the unconditional Sharpe ratio of a buy-and-hold strategy. Assuming the Sharpe ratio of the market portfolio is 0.5, then an R^2 of 40% implies a Sharpe ratio for the market timing strategy that exceeds 1. Such high Sharpe ratios could persist during the financial crisis because of the presence of severe financial constraints that prevent arbitrageurs from taking advantage of the investment opportunities.

Theories of intermediary constraints argue that variations in the constraints not only will affect the risk premium of the market portfolio, but also the risk premia on any financial assets for which the financial intermediaries are the marginal investor. Having examined the ability of $PNBO$ to predict future market excess returns, we now apply the predictability regressions to other asset classes.

Among the assets we consider are (1) high-yield corporate bonds (based on the Barclays U.S. Corporate High Yield total return index), (2) hedge funds (based on the HFRI fund-weighted average return index), (3) carry trade (constructed by Lustig, Roussanov, and Verdelhan, 2011, using the exchange rates of 15 developed countries), (4) commodity (based on the Goldman Sachs commodity index excess return series), (5) the 10-year US Treasury, and (6) variance swap for the S&P 500 index returns (with the excess return defined as the log ratio of the realized annualized return variance over the swap rate; see e.g., Carr and Wu, 2009).[14]

As benchmark, the first row of Table 6 restates the predictability results for equity (market excess returns). It then shows that, our constraint measures predict the future excess returns for a variety of assets besides equity. In periods with $b_{VP,t} < 0$, $PNBO$ predicts negatively and statistically significantly (at least at the 10% level) the future 3-month returns of high yield bonds, hedge funds, carry trade, and commodity (for

[14]Data for the returns on high yield bonds, commodity, and hedge funds are from Datastream. Government bond return data are from Global Financial Data.

Table 6: **Return Forecasts for Various Financial Assets**

This table reports the results of forecasting future excess returns on a variety of assets. All excess returns are in percentages except those for variance swap, which are divided by 100 for scaling. Standard errors in parentheses are computed based on Hodrick (1992). (***,** ,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$r_{t\to t+3} = a_r + b_r^-\ I_{\{b_{VP,t}<0\}}\, PNBO_t + b_r^+\ I_{\{b_{VP,t}\geq 0\}}\, PNBO_t + c_r\ I_{\{b_{VP,t}<0\}} + \epsilon_{t\to t+3}$$

Asset Class	b_r^-	b_r^+	R^2	b_r^-	b_r^+	R^2
	$PNBO$			$PNBON$		
Equity	-84.75***	-42.46**	19.8	-2.67***	-1.58	9.9
	(26.16)	(19.65)		(0.85)	(1.01)	
High Yield	-55.23***	-35.26**	22.2	-1.37***	-1.99***	11.0
	(21.37)	(15.55)		(0.50)	(0.66)	
Hedge Fund	-32.48***	-18.98*	11.6	-1.01***	-0.40	5.4
	(10.67)	(9.82)		(0.33)	(0.44)	
Carry Trade	-38.34**	-23.10	9.1	-0.47	-0.88	2.0
	(18.13)	(14.20)		(0.43)	(0.62)	
Commodity	-71.54*	-43.70	8.1	-1.30	-1.80	2.3
	(42.26)	(30.19)		(0.94)	(1.34)	
10-Year Treasury	20.57**	10.76	5.5	0.79**	0.37	3.8
	(8.92)	(10.65)		(0.33)	(0.45)	
Variance Swap (×1%)	13.64***	4.82**	21.2	0.38***	0.25**	9.5
	(4.02)	(2.16)		(0.11)	(0.11)	

$PNBON$, the coefficient b_r^- for carry trade and commodity returns are still negative but insignificant). Thus, like the market portfolio, the risk premia on these assets tend to rise when the intermediary constraints tighten. The predictive power of $PNBO$ on these asset returns is also economically significant. For example, in a supply environment, a one standard deviation decrease of $PNBO$ is associated with a 2.8% and 1.6% increase in the subsequent 3-month expected excess returns of the high-yield bond index and the hedge fund index, respectively.

Next, both for $PNBO$ and $PNBON$, the predictive coefficient b_r^- is positive and significant for the 3-month excess returns of the 10-year Treasury and the S&P 500 variance swap. Here, a one standard deviation decrease of $PNBO$ in a supply environment is associated with a 1% and 68% decrease in the subsequent 3-month expected excess returns of 10-year Treasuries and variance swaps, respectively. The result on Treasuries is consistent with Fontaine and Garcia (2012), who find the deterioration in funding liquidity predicts lower risk premia for Treasuries. Intuitively, when financial intermediaries become constrained, Treasury values tend to rise ("flight to quality") as does the volatility in the market. Thus, Treasuries and (pay-fix) variance swaps provide a hedge against negative shocks to intermediary constraints.

3.4 An alternative hypothesis

The above return predictability results have two alternative interpretations. It is possible that financial intermediaries become more constrained when the market risk premium rises (e.g., due to higher aggregate uncertainty in the real economy), which in turn reduces their capacity to provide market crash insurance to public investors. As a result, a low $PNBO$ today would be associated with high future market returns, even though a tighter intermediary constraint does not *cause* the market risk premium to rise in this case. Alternatively, it is possible that intermediary constraints directly affect the aggregate market risk premium, which is a central prediction in intermediary asset pricing theories.

To distinguish between these two interpretations, we compare $PNBO$ against a number of financial and macro variables that have been shown to predict market returns. If $PNBO$ is merely correlated with the standard risk factors and does not directly affect the risk premium, then the inclusion of the proper risk factors into the predictability regression should drive away the predictive power of $PNBO$. The variables we consider include the difference between implied and historical volatility used in Bollerslev, Tauchen, and Zhou (2009) ($IVRV$), the log dividend yield ($d - p$) of the market portfolio, the log net payout yield (lcrspnpy) by Boudoukh, Michaely, Richardson, and Roberts (2007), the Baa-Aaa credit spread (DEF), the 10-year minus 3-month Treasury term spread (TERM), the tail risk measure (Tail) by Kelly and Jiang (2014), the slope of the implied volatility curve (IVSlope), and the consumption-wealth ratio measure ($\widehat{cay}$) by Lettau and Ludvigson (2001). All the variables are available monthly except for $\widehat{cay}$, which is at quarterly frequency.

Table 7 shows that, with the inclusion of the various competing variables, the predictive coefficient b_r^- remains significantly negative and remarkably stable in magnitude for both $PNBO$ and $PNBON$. Comparing the R^2 from the regressions in Table 7 and the regression with only $PNBO$ (or $PNBON$) (first row, Table 6), we see that the incremental explanatory power for future market excess returns mostly comes from $PNBO$ ($PNBON$) interacted with the price-quantity relation indicator.[15]

In summary, the results from Table 7 show that the option trading activities of public investors and financial intermediaries contain unique information about the market risk premium that is not captured by the standard macro and financial factors. This result is consistent with the theories of intermediary constraints driving asset prices. Of course, the evidence above does not prove that intermediary constraints actually drive aggregate risk premia. It is possible that $PNBO$ is correlated with other risk factors not considered in our specifications.

3.5 Option volume and measures of funding constraints

We have presented evidence linking $PNBO$ negatively with option expensiveness and market risk premium when $I_{b_{VP,t}<0}$, which is consistent with the interpretation that low $PNBO$ is a sign of tight intermediary constraints. Thus, $I_{b_{VP,t}<0} \times PNBON_t$ can be viewed as a measure of intermediary constraint. We now compare this measure to several measures of financial intermediary funding constraints proposed in the literature.

These measures include the year-over-year change in broker-dealer leverage advocated by Adrian, Moench, and Shin (2010) (Δlev), the fixed-income market based funding liquidity measure by Fontaine and Garcia (2012) (FG), the TED spread (TED, the difference between 3-month LIBOR and 3-month T-bill rate), and the LIBOR-OIS spread (LIBOR-OIS, the difference between 3-month LIBOR and 3-month overnight indexed swap

[15]In theory, variation in risk premium due to intermediary constraints should affect variables such as the dividend-price ratio. The reason that $d - p$ does not show significant predictive power in Table 7 in the presence of $PNBO$ could be that $d - p$ is affected by both the variations in discount rates and expected dividend growth, and that transitory fluctuations in the discount rate caused by fluctuations in intermediary constraints have limited effects on prices.

Table 7: **Return Forecasts with $PNBO$ and Other Predictors**

The table reports the results of forecasting 3-month market excess returns with $PNBO$ and other predictors, including the difference between implied and historical volatility in Bollerslev, Tauchen, and Zhou (2009) (IVRV), log dividend yield ($d-p$), log net payout yield (lcrspnpy), Baa-Aaa credit spread (DEF), term spread (TERM), tail risk measure (Tail), implied volatility slope (Slope) and the quarterly consumption-wealth ratio ($\widehat{cay}$). Standard errors (in parentheses) are computed based on Hodrick (1992). (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991 – 2012, except for lcrspnpy and Tail (1991 – 2010), and IVSlope (1996 – 2012).

$$r_{t\to t+3} = a_r + b_r^- \; I_{\{b_{VP,t}<0\}} \, PNBO_t + b_r^+ \; I_{\{b_{VP,t}\geq 0\}} \, PNBO_t + c_r \; I_{\{b_{VP,t}<0\}} + d_r \; \mathbf{X}_t + \epsilon_{t\to t+3}$$

Panel A: $PNBO$

$I_{\{b_{VP,t}<0\}} \, PNBO_t$	-77.22***	-82.44***	-94.46***	-85.81***	-87.33***	-90.86***	-75.80***	-80.13***	-62.70***
	(26.09)	(26.66)	(30.95)	(26.61)	(26.96)	(31.35)	(24.59)	(31.40)	(22.71)
$I_{\{b_{VP,t}\geq 0\}} \, PNBO_t$	-18.50	-37.81*	-47.38**	-44.73**	-46.26**	-47.54**	-36.40*	-14.42	-59.22
	(20.16)	(20.21)	(22.76)	(21.02)	(20.32)	(22.94)	(19.84)	(27.59)	(37.45)
IVRV	0.10***							0.13***	
	(0.04)							(0.04)	
$d-p$		3.54						11.39	
		(3.20)						(7.75)	
lcrspnpy			6.19					7.75	
			(4.89)					(8.10)	
DEF				-1.58				-5.96	
				(3.03)				(3.83)	
Term					-0.63			-0.63	
					(0.71)			(0.83)	
Tail						30.56		-52.06	
						(35.02)		(47.76)	
IVSlope							0.29	0.17	
							(0.26)	(0.29)	
$\widehat{cay}$									92.71**
									(46.30)
R^2	25.2	21.3	22.9	20.4	20.6	20.8	22.4	39.2	19.9

Panel B: $PNBON$

$I_{\{b_{VP,t}<0\}} \, PNBON_t$	-2.50***	-2.42***	-2.47***	-2.70***	-2.65***	-2.56***	-2.63***	-2.06***	-2.07**
	(0.85)	(0.85)	(0.89)	(0.86)	(0.85)	(0.89)	(1.00)	(1.08)	(0.86)
$I_{\{b_{VP,t}\geq 0\}} \, PNBON_t$	-0.85	-1.20	-1.41	-1.78	-1.61	-1.60	-1.17	0.06	-4.81*
	(1.01)	(1.05)	(1.12)	(1.08)	(1.02)	(1.11)	(1.08)	(1.32)	(2.60)
IVRV	0.13***							0.15***	
	(0.04)							(0.04)	
$d-p$		3.13						13.63	
		(3.17)						(7.42)	
lcrspnpy			3.13					1.51	
			(5.02)					(7.92)	
DEF				-1.71				-5.57	
				(3.03)				(3.89)	
Term					-0.27			-0.10	
					(0.67)			(0.85)	
Tail						26.94		-38.27	
						(34.87)		(46.84)	
IVSlope							0.36	0.33	
							(0.28)	(0.31)	
$\widehat{cay}$									108.03**
									(46.11)
R^2	19.2	10.7	9.7	10.3	9.7	9.7	14.5	32.5	17.7

rate).[16] TED spread and LIBOR-OIS spread measure the credit risk of banks.

We first run OLS regressions of our constraint measure on the funding constraint measures in the literature. As Panel A of Table 8 shows, TED spread is significantly positively related to $I_{b_{VP,t}<0} \times PNBO_t$, but insignificantly related to $I_{b_{VP,t}<0} \times PNBON_t$. The positive relation between TED spread and $I_{b_{VP,t}<0} \times PNBO_t$ is mainly due to the fact that $PNBO$ rose significantly along with the TED spread during the early part of the financial crisis. Subsequently, while $PNBO$ becomes lower (and turned significantly negative), the TED spread also fell and then remained at low levels. This result points out a potential weakness of TED spread (and LIBOR-OIS) as a measure of funding constraint. TED spread could become lower due to banks reducing their own credit risk through deleveraging, reducing risk taking, and buying crash insurance, but that does not necessarily imply banks are less constrained (could be the opposite).

Next, our constraint measure is significantly negatively related to the measure FG. This is consistent with FG's interpretation that financial intermediaries are more constrained (low $PNBO$) in periods when the value of funding liquidity is high (high FG). In the quarterly regression, our constraint measure is significantly positively related to the growth rate in broker-dealer leverage Δlev. That is, intermediary constraint tends to be tight (low $PNBO$) when broker-dealers are de-leveraging (low Δlev).[17]

In Panel B of Table 8, we further examine the ability of the various funding constraint measures to predict aggregate market returns. Adrian, Moench, and Shin (2010) show that Δlev has strong predictive power for excess returns on stocks, corporate bonds, and treasuries. In a univariate regression (unreported) with Δlev, we find similar results in our sample period. When joint with our constraint measure, the coefficient on Δlev becomes insignificant in the case of $PNBO$ and marginally significant in the case of $PNBON$, while b_r^- remains significant. Similarly, when the other funding constraint measures are used in place of Δlev, the coefficient b_r^- is always highly significant. These results suggest that relative to other funding constraint measures, our intermediary constraint measure

[16] We also examine the funding constraint measure by Hu, Pan, and Wang (2013) and the CBOE VIX index (VIX). Neither of them is statistically significantly related to PNBO.

[17] He, Kelly, and Manela (2017) show that the leverage of commercial banks became higher during the financial crisis while that of broker-dealers fell.

Table 8: $PNBO$ **and Measures of Funding Constraints**

Panel A reports the results of the OLS regressions of $I_{b_{VP,t}<0} \times PNBO_t$ on measures of funding constraints. Panel B reports the results of return predictability regressions with $PNBO$ and funding constraint measures. TED is the TED spread; LIBOR-OIS is the spread between 3-month LIBOR and overnight indexed swap rates; FG is the funding liquidity measure by Fontaine and Garcia (2012); Δlev is the broker-dealer balance sheet growth measure by Adrian, Moench, and Shin (2010). Standard errors (in parentheses) are computed based on Newey and West (1987) with 3 lags. (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991 – 2012, except for the regressions with LIBOR-OIS (2002 – 2012).

Panel A: $I_{b_{VP,t}<0} \times PNBO_t = a + b\ \mathbf{X}_t + e_t$

	$PNBO$				$PNBON$			
TED	22.04**				0.56			
	(10.57)				(147.30)			
LIBOR-OIS		0.25				-0.17		
		(0.16)				(2.04)		
FG			-4.89**				-200.65***	
			(2.32)				(77.30)	
Δlev				42.59**				50.86**
				(16.68)				(23.18)
R^2	4.3	2.9	1.5	8.9	0.0	0.0	4.4	3.4

Panel B: $r_{t\to t+3} = a_r + b_r^-\ I_{\{b_{VP,t}<0\}}\ PNBO_t + b_r^+\ I_{\{b_{VP,t}\geq 0\}}\ PNBO_t + c_r\ I_{b_{VP,t}<0} + d_r\ \mathbf{X}_t + \epsilon_{t\to t+3}$

	$PNBO$				$PNBON$			
$I_{\{b_{VP,t}<0\}}\ PNBO_t$	-79.97***	-75.91***	-83.52***	-53.78**	-2.62***	-3.81***	-2.57***	-1.63**
	(24.66)	(23.94)	(26.65)	(22.47)	(0.84)	(1.23)	(0.85)	(0.83)
$I_{\{b_{VP,t}\geq 0\}}\ PNBO_t$	-34.84*	-38.41**	-42.00**	-37.02	-1.38	-3.07**	-1.50	-3.69
	(19.34)	(19.37)	(19.65)	(40.89)	(1.02)	(1.27)	(1.01)	(2.79)
TED	-2.44				-4.83			
	(3.19)				(3.25)			
LIBOR-OIS		-0.05				-0.08		
		(0.05)				(0.05)		
FG			0.50				0.24	
			(0.88)				(0.86)	
ΔLev				-4.76				-5.82*
				(3.51)				(3.41)
R^2	20.4	34.9	19.6	19.9	14.4	33.1	9.6	18.9

contains unique information about conditional market risk premia.

3.6 Econometric identification of supply shocks

Since equilibrium prices and quantities are jointly determined by supply and demand, a key element in establishing a causal link between supply and risk premia would be a method of identifying exogenous shocks to supply. Rigobon (2003) solves the identification problem in the supply and demand equations (4a–4b) through heteroscedasticity. We implement this method using VP as the price variable and $PNBO$ (or $PNBON$) as the quantity variable. Following Rigobon (2003) (who uses crisis periods in debt markets to identify movement in bond markets in Latin America), we consider the high supply volatility (relative to demand volatility) regime for our identification as the period from September 2008 (the Lehman default) to November 2011 (cancellation of Greek referendum). The second regime is the remainder of our sample.

The results of the estimation are shown in Table 9. We see when using either $PNBO$ or $PNBON$ as the quantity measure, there is a statistically significant positive (negative) slope for the supply (demand) curve. The point estimates indicate that a one-standard deviation increase in the variance premium (an increase of 21.7) is associated with an increased quantity of 103,000 deep out-of-the-money put contracts, which represents an increase of 2.02 standard deviations to the equilibrium quantity. In terms of the normalized measure $PNBON$, a one-standard deviation increase in the variance premium results in an increased quantity of 5.8% times the trailing 3-month volume of SPX options.

Table 10 reproduces the predictability results of Table 4 replacing $PNBO$ and $PNBON$ with the extracted supply shocks and normalized supply shocks.[18] The extracted supply shock is strongly related to $PNBO$ with a correlation of 0.71. Overall, the predictability with the extracted shocks is very similar to what we found with $PNBO$ and $PNBON$, which reinforces our interpretation of the results.

However, we temper this evidence with a caveat based on the assumptions of the

[18]In the regression, we use the Hodrick (1992) standard errors to be consistent with the remainder of the paper, which do not correct for the error in variables associated with uncertainty in the parameters used to extract the supply shocks.

Table 9: **Supply-demand estimation**

This table reports parameter estimates from estimation of the supply-demand system given by

$$\text{Demand: } VP_t = b + \beta \cdot PNBO_t + \epsilon_t,$$
$$\text{Supply: } PNBO_t = a + \alpha \cdot VP_t + \eta_t,$$

using the econometric identification of Rigobon (2003). The η and ϵ are uncorrelated with regime-dependent volatilities. We set the high supply volatility regime to be from September 2009 to November 2011. Standard errors are computed by bootstrap. Sample period: 1991/01 – 2012/12. (***,**,*) denote significance at 1%, 5%, and 10%, respectively; 1-sided p-values are computed for α and β.

	$PNBO$		$PNBON$	
b	25.9^{***}	(3.96)	55.5^{*}	(34.8)
β	-0.0823^{*}	(0.0541)	-4.63^{**}	(2.37)
a	-655^{*}	(340)	-63.4	(53.5)
α	0.00477^{**}	(0.00276)	0.269^{**}	(0.12)

methodology and its application to our setting. The method used here to identify supply shocks assumes linear and stationary supply/demand relationships, and zero correlation between supply and demand shocks.[19] The results are likely sensitive to these assumptions.

4 Robustness Checks

In this section, we report the results of several robustness checks for our main results.

4.1 Financial crisis

One potential concern regarding the predictive power of $PNBO$ is that it might be driven by a small number of outliers, in particular the 2008-09 financial crisis. To address this concern, Table 11 reports the results of the return-forecasting regressions in two sub-samples: pre-crisis (1991/01-2007/11) and post-crisis (2009/06-2012/12). The predictive powers of $PNBO$ and $PNBON$ remain statistically significant in both sub-samples. The

[19] One could easily imagine non-linear relationships (such as a dependency on the level of jump risk) or slope coefficients that vary with the volatility regime, as well as non-zero correlation between supply and demand shocks due to exogenous factors simultaneously driving supply and demand.

Table 10: **Return Forecasts with supply shocks**

This table reports the results of the return forecasting regressions using supply shocks extracted using the econometric identification of Rigobon (2003). We set the high supply volatility regime to be from September 2009 to November 2011. $r_{t+j\to t+k}$ represents log market excess return from month $t+j$ to $t+k$ ($k > j \geq 0$). Standard errors (in parentheses) are computed based on Hodrick (1992). (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$r_{t+j\to t+k} = a_s + b_s\hat{\eta}_t + \epsilon_{t+j\to t+k}$$

Horizon	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2
	PNBO supply shocks			PNBON supply shocks		
$r_{t\to t+1}$	-5.80*	(3.25)	2.7	-0.19**	(0.08)	2.8
$r_{t+1\to t+2}$	-5.23	(3.30)	2.2	-0.14**	(0.07)	1.7
$r_{t+2\to t+3}$	-7.92***	(2.83)	5.1	-0.06	(0.06)	0.3
$r_{t+3\to t+4}$	-6.32**	(3.02)	3.2	-0.02	(0.06)	0.0
$r_{t\to t+3}$	-18.80**	(7.48)	8.8	-0.39**	(0.16)	3.8

economic significance of the predictive power is weaker in the pre-crisis period than in the full sample (in terms of smaller magnitude of b_r^- and lower R^2), but it is quite strong in the post-crisis period. Thus, the predictive power of our measure of intermediary constraint is not just a crisis phenomenon. The weaker predictive power for $PNBO$ in the earlier sample period could be due to the fact that intermediary constraints are not as significant and volatile in the first half of the sample. Another reason might be that the SPX options market was less developed in the early periods and did not play as important a role in facilitating risk sharing as it does today.

4.2 Moneyness

Next, we examine how the predictive power of $PNBO$ changes based on option moneyness. Our baseline definition of DOTM puts uses a simple cutoff rule $K/S \leq 0.85$. Panel A of Figure 4 plots the coefficient b_r^- and the confidence intervals as we change this cutoff value for SPX puts. The coefficient b_r^- in the return forecast regression is significantly negative for a wide range of moneyness cutoffs. The point estimate of b_r^- does become more negative as the cutoff becomes smaller. Because DOTM puts are more difficult to

Table 11: **Return Forecasting Outside the Financial Crisis**

This table reports the 3-month return forecasting results outside the financial crisis. Standard errors (in parentheses) are computed based on Hodrick (1992). (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Pre-crisis: 1991/01 - 2007/11. Post-crisis: 2009/06 - 2012/12.

$$r_{t\to t+3} = a_r + b_r^- \; I_{\{b_{VP,t}<0\}}\, PNBO_t + b_r^+ \; I_{\{b_{VP,t}\geq 0\}}\, PNBO_t + c_r \; I_{\{b_{VP,t}<0\}} + \epsilon_{t\to t+3}$$

	b_r^-	b_r^+	R^2	b_r^-	b_r^+	R^2
	$PNBO$			$PNBON$		
Pre-crisis	-66.50**	21.91	3.8	-1.85**	-0.12	5.3
	(27.53)	(22.90)		(0.92)	(1.12)	
Post-crisis	-50.88**	-15.73	17.4	-4.87**	-1.02	19.6
	(25.72)	(33.42)		(2.39)	(3.61)	

hedge than ATM puts, they expose financial intermediaries to more inventory risks. Hence, $PNBO$ measure based on DOTM puts should be more informative about intermediary constraints and in turn the aggregate risk premium than $PNBO$ based on ATM puts. At the same time, because far out-of-money options are less liquid, the $PNBO$ series becomes more noisy when we further reduce the cutoff, which widens the confidence interval on b_r^-. In contrast, Panel B shows that for essentially all moneyness cutoffs, $PNBO$ based on SPX call options does not predict market returns.

A feature of our definition of DOTM puts above is that a constant strike-to-price cutoff implies different actual moneyness (e.g., as measured by option delta) for options with different maturities. A 15% drop in price over one month might seem very extreme in calm periods, but it is more likely when market volatility is high. For this reason, we also examine a maturity-adjusted moneyness definition. Specifically, we classify a put option as DOTM when $K/S \leq 1 + k\sigma_t\sqrt{T}$, where k is a constant, σ_t is the daily S&P return volatility in the previous 30 trading days, and T is the days to maturity for the option. This is similar to using option delta to define moneyness, but does not require a particular pricing model to compute the delta. Panel C of Figure 4 shows that this alternative classification of DOTM puts produces qualitatively similar results as our simple cutoff rule. Panel D shows again that $PNBO$ based on SPX calls and this alternative

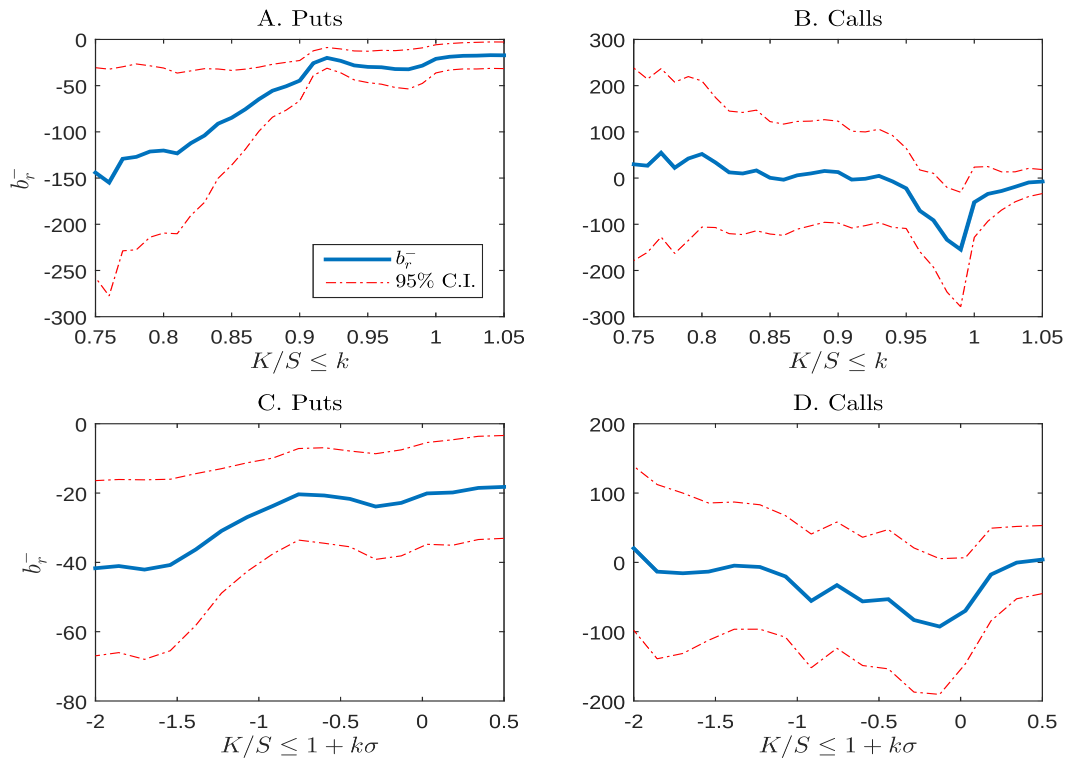

Figure 4: **Predictive power of $PNBO$ at different moneyness.** This figure plots the point estimate and 95% confidence interval (based on Hodrick (1992) standard errors) for the coefficient b_r^- from regression (5) forecasting 3-month market excess returns. In Panels A and B, option moneyness is measured by the level of strike-to-price ratio K/S. In Panels C and D, option moneyness is measured by K/S relative to daily volatility of S&P returns scaled by the square root of days to maturity (see Section 4.2).

moneyness cutoff does not predict returns.

4.3 Volume vs. open interest

In our construction of $PNBO$, we focus on the net amount of new DOTM index puts that public investors buy in a period. An alternative way to gauge the economic exposure for public investors and financial intermediaries is to examine the net open interest for the two groups. Between new volume and open interest, which one better represents the degree of intermediary constraints?

We use the volume-based measure in our main analysis for a few reasons. The first reason is data limitation, as CBOE does not provide daily long and short open interest data after 2001. Second, when financial intermediaries become constrained, it is likely easier (e.g., due to transaction costs) to adjust the quantity of new DOTM puts traded than to change their established positions. That would make the volume-based measure more sensitive to changes in intermediary constraints than open interest-based measure. Third, taking on an option position that is originally near the money but later becomes DOTM due to market movements is different from taking on a new DOTM option. In the former case, the intermediaries can put on hedges against tail risk over time, which again means these positions are less sensitive to changes in intermediary constraints.

Nonetheless, we provide two robustness checks. First, we examine an alternative measure based on the end-of-month public net open interest for DOTM SPX puts ($PNOI$). The results are discussed below. Second, we construct a $PNBO$ measure using only one-month options, for which the monthly measures of net volume and open interest are equivalent. We find that the main results hold for the measure based on short-dated options (see Table IA4 in the Internet Appendix).

For the period of 1991 to 2001, we use daily long and short open interest provided by CBOE to compute $PNOI$. From 2001 onward, we compute daily net open interest from the volume information as follows:

$$NOI_d^{K,T} = NOI_{d-1}^{K,T} + \mathrm{openBuy}_\mathrm{d}^\mathrm{K,T} - \mathrm{openSell}_\mathrm{d}^\mathrm{K,T} + \mathrm{closeBuy}_\mathrm{d}^\mathrm{K,T} - \mathrm{closeSell}_\mathrm{d}^\mathrm{K,T}, \tag{7}$$

where $NOI_d^{K,T}$ is the public investor net open interest of options with strike price K and maturity T on day d, openBuy is the public investor buying volume from initiating long positions, openSell is the selling volume from initiating short, closeBuy is the buying volume from closing existing short positions, and closeSell is the selling volume from closing existing long positions. We then aggregate $NOI_d^{K,T}$ to compute daily net open interest of DOTM puts ($PNOI$). We also consider a normalized version of $PNOI$ ($PNOIN$), which is $PNOI$ divided by the sum of public long and short open interest for all SPX options.

Table 1 shows that the end-of-month $PNOI$ is around 29,000 contracts on average, and $PNOI$ has higher autocorrelation than $PNBO$. Table 12 shows that, like $PNBO$, $PNOI$ predicts future market excess returns negatively in the periods with $b_{VP,t} < 0$, with

Table 12: **Return Forecasts with $PNOI$**

This table reports the results of the return forecasting regressions using $PNOI$ and $PNOIN$. Standard errors (in parentheses) are computed based on Hodrick (1992). (***,**,*) denote significance at 1%, 5%, and 10%, respectively. Sample period: 1991/01 – 2012/12.

$$r_{t+j\to t+k} = a_r + b_r^- I_{\{b_{VP,t}<0\}}\, PNOI_t + b_r^+ I_{\{b_{VP,t}\geq 0\}}\, PNOI_t + c_r I_{\{b_{VP,t}<0\}} + \epsilon_{t+j\to t+k}$$

Horizon	b_r^-	b_r^+	R^2	b_r^-	b_r^+	R^2
	$PNOI$			$PNOIN$		
$r_{t\to t+1}$	-25.10***	-11.72	8.1	-3.71***	-1.23	5.6
	(8.65)	(7.93)		(1.11)	(1.17)	
$r_{t+1\to t+2}$	-19.79**	-1.65	5.3	-2.11*	-0.25	2.7
	(9.59)	(4.72)		(1.17)	(0.99)	
$r_{t+2\to t+3}$	-19.34**	-4.67	4.5	-2.21**	-0.26	2.0
	(7.73)	(7.66)		(0.97)	(1.19)	
$r_{t+3\to t+4}$	-9.75	-17.17*	4.2	-1.32	-2.32*	2.4
	(6.85)	(9.18)		(0.96)	(1.29)	
$r_{t\to t+3}$	-64.22***	-18.05	14.7	-8.03***	-1.74	7.8
	(23.47)	(13.50)		(2.68)	(2.28)	

significant coefficient b_r^- up to 3 months in the future. The R^2 of the regressions are also similar to $PNBO$. Additional sub-sample results for $PNOI$ are provided in Table IA5.

4.4 Public investors: retail vs. institutional

As Figure 1 shows, while financial intermediaries typically sell DOTM SPX puts to public investors during normal times, the roles are often reversed during crisis times, most notably during the 2008-09 financial crisis. To understand the risk sharing mechanism between financial intermediaries and public investors, it is informative to find out who among the public investors are the "liquidity providers," reducing the demand for crash insurance or even providing insurance to the intermediaries when the latter become constrained. The SPX volume data from CBOE do not provide further information about the types of public investors behind a given transaction (e.g., retail vs. institutional investors). We tackle this question by comparing the trading activities of the public investors in SPX options with those in SPY options.

While SPX and SPY options have essentially identical underlying asset, it is well known among practitioners that institutional investors account for a significantly higher percentage of the trading volume of SPX options than do SPY options. Compared to retail investors, institutional investors prefer SPX options due to a larger contract size (10 times as large as SPY), cash settlement, more favorable tax treatment, as well as being more capable of trading in between the relatively wide bid-ask spreads of SPX options due to stronger bargaining power. We construct $PNBO_{SPY}$ for SPY options using the same procedure as $PNBO$, which covers the period from 2005/05 to 2012/12. While SPX options trade exclusively on the CBOE, SPY options are cross-listed at several option exchanges. $PNBO_{SPY}$ aggregates the volume from the CBOE and International Securities Exchange (ISE), which account for about half of the total trading volume for SPY options.

Figure 5 compares the $PNBO_{SPY}$ and $PNBO_{SPX}$ (equivalent to $PNBO$) series. During the period of 2005/05 to 2012/01, $PNBO_{SPY}$ is positive in the majority of the months. From 2008/09 to 2010/12, $PNBO_{SPX}$ is negative in 22 out of 28 months, whereas $PNBO_{SPY}$ is negative in just 7 of the 28 months.

A systematic way to examine the difference in how the equilibrium quantities of trading in the two markets are connected to the intermediary constraints and market risk premium is through the regressions of (3) and (5). These results are reported in the Online Appendix, and we summarize the main findings here. Unlike SPX, the $PNBO$ measure based on SPY is insignificantly (positively) related to the variance premium on average, and does not predict future market excess returns. The contrast between SPX and SPY suggests that institutional investors and retail investors respond differently to the changes in intermediary constraints. In particular, when constrained financial intermediaries start buying crash insurance, they appear to be buying the insurance from the public (institutional) investors in the SPX market and not from the public (retail) investors in the SPY market.[20]

[20]Notice that this difference in public investor trading behaviors between the two markets does not necessarily imply arbitrage.

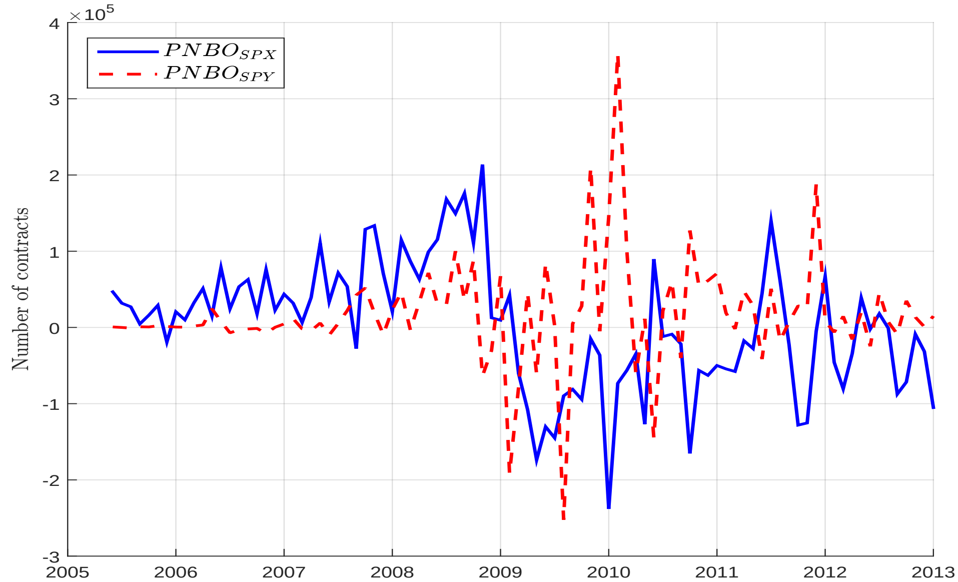

Figure 5: **Comparing $PNBO$ for SPX and SPY options.** This figure plots the public net buying-to-open volume for deep-out-of-the-money (with $K/S \leq 0.85$) puts in the market for SPX options and SPY options for the period of 2005/05 to 2012/12.

4.5 Comparison with GPP

Our results on the price-quantity relation in the DOTM SPX puts market is related to Garleanu, Pedersen, and Poteshman (2009). However, our results differ from GPP in several aspects. First, we have a longer sample period, from 1991 to 2012, while theirs is from 1996 to 2001. Second, our $PNBO$ measure uses contemporaneous public net-buying volume, while GPP use public net open interest,[21] which is the accumulation of past net-buying volumes. Third, our $PNBO$ measure focuses on DOTM SPX puts, whereas GPP use options of all moneyness (which is similar to the sum of $PNBO$ and $PNBO_{ND}$ in this regard). In this section, we first replicate the main results of GPP (in Table 2, p4287), and then examine the differences between the two studies.

The dependent variable in Table 13, option expensiveness, is the same as the one used in GPP Table 2, i.e., the average implied volatility of ATM options minus a reference model-implied volatility used in Bates (2006).[22] GPP regress the option expensiveness

[21] GPP aggregate the net demand of both public investors and firm investors, and refer to it as the non-market-maker demand.

[22] We thank David Bates for sharing the data on this measure.

Table 13: **Comparison with GPP Table 2**

The table reports the results from regressing option expensiveness, measured by the average implied volatility of ATM options minus a reference model-implied volatility used in Bates (2006), on measures of option demand. NetDemand and JumpRisk are the equal- and vega-weighted public net open interest (or net volume) for all SPX options. t-stats are computed based on Newey and West (1987) standard errors with 10 lags. (***,** ,*) denote significance at 1%, 5%, and 10%, respectively.

	Net Open Interest				Net Volume	
	Replicated Results		GPP Results			
	NetDemand	JumpRisk	NetDemand	JumpRisk	NetDemand	JumpRisk
1996/01-1996/10	2.1×10^{-7}	4.5×10^{-6}	2.1×10^{-7}	6.4×10^{-6}	1.6×10^{-8}	-5.1×10^{-6}
t-stat	(0.72)	(0.49)	(0.87)	(0.79)	(0.10)	(-0.23)
1997/10-2001/12	$3.8\times10^{-7**}$	$2.6\times10^{-5***}$	3.8×10^{-7}	$3.2\times10^{-5***}$	$4.7\times10^{-6**}$	$4.9\times10^{-5**}$
t-stat	(2.17)	(2.80)	(1.55)	(3.68)	(2.25)	(2.57)
2002/01-2012/12	$-9.9\times10^{-8***}$	$-7.8\times10^{-6**}$			$-6.7\times10^{-7**}$	$-5.2\times10^{-5***}$
t-stat	(-3.13)	(-2.40)			(-2.35)	(-3.32)

on several measures of SPX non-market-maker demand pressure, including the equal-weighted public open interest (NetDemand), and public open interest weighted by jump risk (JumpRisk). They find the regression coefficients to be positive in the period from 1997/10 to 2001/12.

In Table 13, we obtain very similar results to GPP for the two sub-samples 1996/01-1996/10 and 1997/10-2001/12 for the open interest-based measures. We also construct two net volume-based measures and again find that they are positively related to option expensiveness in the period from 1997/10 to 2001/12.

Next, for the period 2002-2012, we use daily volume data to extend the net open interest measures (constructed using the procedure described in Section 4.3). In this subsample, We find that the coefficients on both the open interest and volume-based measures become negative and statistically significant. This finding is consistent with our finding of a negative price-quantity relation in the full sample (see Table 3). The changing signs of the price-quantity relation in different sub-samples suggest that the effects of demand shocks and supply shocks are both present in the SPX option market.

5 Conclusion

We provide evidence that the trading activities of financial intermediaries in the market of DOTM SPX put options are informative about the degree of intermediary constraints. In periods when supply shocks are likely to be the main force behind the variations in the price-quantity relation in the DOTM SPX put market, our public investor net-buying volume measure, $PNBO$, has strong predictive power for future market excess returns and the returns for a range of other financial assets. The predictive power of $PNBO$ is stronger during periods when the market jump risk is high, and it is stronger for DOTM puts. $PNBO$ is also associated with several funding liquidity measures in the literature. Moreover, the information that $PNBO$ contains about the market risk premium is not captured by the standard financial and macro variables. These results suggest that time-varying intermediary constraints are driving the supply of crash insurance by financial intermediaries and the risk premia in financial markets.

Appendix

A Physical Jump Risk

The construction of the jump risk measure J follows Bekaert and Hoerova (2014) and Corsi, Pirino, and Reno (2010), which we briefly outline here. The monthly jump J is the average of daily jump J^d, which is defined as

$$J_t^d = \max[RV_t^d - TBPV_t^d, 0], \tag{A1}$$

where RV_t^d is the S&P500 daily realized variance (based on 5-minute returns), and $TBPV_t^d$ stands for threshold bipower variation, which is defined as follows (see Corsi, Pirino, and Reno, 2010, equation (2.14)):

$$TBPV_t^d = \frac{1}{0.7979^2} \sum_{j=2}^{N} |r_{j-1} \cdot r_j| \cdot I_{r_{j-1}^2 \leq \vartheta_{j-1}} \cdot I_{r_j^2 \leq \vartheta_j}. \tag{A2}$$

Here, r_j is the jth 5-minute return on day t; ϑ_j is a stochastic threshold,

$$\vartheta_j = c_\vartheta^2 \cdot \hat{V}_j, \tag{A3}$$

where c_ϑ is a scale-free constant (set to 3), and $\hat{V}_j$ is an auxiliary estimator of variance, which is estimated through the following iteration,

$$\hat{V}_j^Z = \frac{\sum_{i=-L, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) r_{j+i}^2 I_{\left\{r_{j+i}^2 \leq c_\vartheta^2 \cdot \hat{V}_{j+i}^{Z-1}\right\}}}{\sum_{i=-L, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) I_{\left\{r_{j+i}^2 \leq c_\vartheta^2 \cdot \hat{V}_{j+i}^{Z-1}\right\}}}, \quad Z = 1, 2, ... \tag{A4}$$

The starting value is $\hat{V}_j^0 = +\infty$. In each iteration step, large returns are eliminated based on the condition $r_j^2 > c_\vartheta^2 \cdot \hat{V}_j^{Z-1}$, and the iteration stops when there are no more large returns to remove. The bandwidth parameter L determines the number of adjacent returns included in the estimation of the local variance around point j (with the observation at j and the two adjacent ones excluded). We set $L = 25$ and use a Gaussian kernel $K(y) = (1/\sqrt{2\pi}) \exp(-y^2/2)$.

B Additional Empirical Results

To check whether our supply-environment indicator $I_{b_{VP,t}<0}$ does a good job identifying important events of supply shocks, Table A1 shows the value of our indicator during a set of months in which significant events have occurred in the financial markets, and which are likely associated with either positive or negative supply shocks for financial intermediaries.

In addition, Table A1 also reports the value of the indicators based on sign of $b_{VP,t}$ and the CDM method from November 2007 to June 2009, the period of the 2008 financial crisis. Outside the 2008 crisis, the strategy based on the sign of $b_{VP,t}$ identifies the major events seven times out of ten, whereas the CDM strategy identifies two events. Inside the 2008 crisis, the two indicators both pick up 12 out of 20 months.

Additional empirical results and robustness checks are provided in the Internet Appendix. They include (1) a table of correlations between $PNBO$ and various macroeconomic and financial variables (Table IA2); (2) systematic analysis of the statistical significance for the return-forecasting regressions using $PBNO$ and $PNBON$ (Table IA3); (3) return-forecasting regression with $PNBO_{1month}$ (and $PNBON_{1month}$), which is $PNBO$ constructed using only options with one month or less to maturity (Table IA4); (4) sub-sample return-forecasting regressions with $PNOI$ (Table IA5); (5) return-forecasting regression with PNB (and $PNBN$), which is the public net buy volume for DOTM SPX puts including both open and close transactions (Table IA6); (6) return-forecasting regression using a modified CDM method (Table IA7); (7) return-forecasting regression using $b_{VP} + c_{VP}J_t$ as indicator (Table IA8); (8) sub-sample return-forecasting regressions with the log dividend-price ratio (Table IA9).

References

Adrian, Tobias, and Nina Boyarchenko, 2012, Intermediary leverage cycles and financial stability, Staff reports, Federal Reserve Bank of New York.

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Adrian, Tobias, Emanuel Moench, and Hyun Song Shin, 2010, Financial intermediation, asset prices, and macroeconomic dynamics, Staff Reports 422 Federal Reserve Bank of New York.

Andersen, T. G., T. Bollerslev, and F. X. Diebold, 2007, Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.

Andrews, Donald W. K., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821–856.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Bates, David S., 2006, Maximum Likelihood Estimation of Latent Affine Processes, *Review of Financial Studies* 19, 909–965.

——— , 2008, The market for crash risk, *Journal of Economic Dynamics and Control* 32.

Bekaert, Geert, and Marie Hoerova, 2014, The VIX, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.

Bernanke, Ben, and Mark Gertler, 1989, Agency Costs, Net Worth, and Business Fluctuations, *American Economic Review* 79, 14–31.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, 1999, The financial accelerator in a quantitative business cycle framework, in J. B. Taylor, and M. Woodford, ed.: *Handbook of Macroeconomics*, vol. 1 of *Handbook of Macroeconomics* . chap. 21, pp. 1341–1393 (Elsevier).

Bollen, Nicolas, and Robert Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts, 2007, On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing, *Journal of Finance* 62, 877–915.

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.

Brunnermeier, Markus K., and Yuliy Sannikov, 2014, A Macroeconomic Model with a Financial Sector, *American Economic Review* 104, 379–421.

Buraschi, Andrea, and Alexei Jiltsov, 2006, Model uncertainty and option markets with heterogeneous beliefs, *Journal of Finance* 61, 2841–2897.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.

Chen, Hui, Scott Joslin, and Ngoc-Khanh Tran, 2012, Rare disasters and risk sharing with heterogeneous beliefs, *Review of Financial Studies* 25, 2189–2224.

Cochrane, John H., 1999, Portfolio advice of a multifactor world, *Economic Perspectives* pp. 59–78.

Cohen, Lauren, Karl B. Diether, and Christopher J. Malloy, 2007, Supply and Demand Shifts in the Shorting Market, *Journal of Finance* 62, 2061–2096.

Corsi, Fulvio, Davide Pirino, and Roberto Reno, 2010, Threshold bipower variation and the impact of jumps on volatility forecasting, *Journal of Econometrics* 159, 276–288.

Fontaine, Jean-Sebastien, and Rene Garcia, 2012, Bond Liquidity Premia, *Review of Financial Studies* 25, 1207–1254.

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.

Geanakoplos, John, 2009, The leverage cycle, Cowles Foundation Discussion Papers 1715 Cowles Foundation for Research in Economics, Yale University.

Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361–407.

Hansen, P.R., and A. Timmermann, 2012, Choice of sample split in out-of-sample forecast evaluation, Working Paper, Stanford University.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics,* forthcoming.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary Asset Pricing, *American Economic Review* 103, 732–770.

Hodrick, RJ, 1992, Dividend yields and expected stock returns: alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2013, Noise as Information for Illiquidity, *Journal of Finance* 68, 2341–2382.

Johnson, Timothy C., Mo Liang, and Yun Liu, 2016, What drives index options exposure?, *Review of Finance*.

Kelly, Bryan, and Hao Jiang, 2014, Editor's choice tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105, 211–48.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Longstaff, Francis A., and Jiang Wang, 2012, Asset pricing and the credit market, *Review of Financial Studies* 25, 3169–3215.

Lustig, Hanno, Nikolai Roussanov, and Adrien Verdelhan, 2011, Common Risk Factors in Currency Markets, *Review of Financial Studies* 24, 3731–3777.

McCracken, Michael W., 2007, Asymptotics for out of sample tests of Granger causality, *Journal of Econometrics* 140, 719–752.

Muller, Ulrich K., 2014, Hac corrections for strongly autocorrelated time series, *Journal of Business & Economic Statistics* 32, 311–322.

Newey, W. K., and K. D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 55, 703–708.

Pan, Jun, and Allen Poteshman, 2006, The information in option volume for future stock prices, *Review of Financial Studies* 19, 871–908.

Rigobon, R, 2003, Identification through heteroskedasticity, *Review of Economics and Statistics* 85, 777–792.

Sentana, Enrique, and Gabriele Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics* 102, 143–164.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Table A1: **Indicators of Supply Environment**

This table reports the indicators of supply environment based on the sign of $b_{VP,t}$ and based on the Cohen, Diether, and Malloy (2007) (CDM) method. The months listed include those associated with major events in the global financial markets. They also include the period of 2007/11 to 2009/06, the period of the 2008 financial crisis. For more details of the events during the 2008 financial crisis, see the Federal Reserve Bank of St. Louis webpage (https://www.stlouisfed.org/financial-crisis/full-timeline).

Month	$I_{b_{VP,t}<0}$	CDM	Events
1997/10	0	0	Asian financial crisis
1998/11	1	0	Russian financial crisis
2003/04	1	0	Start of the Iraq War
2007/08	0	1	Quant crisis
2007/11	1	1	Liquidity deterioration in interbank funding market
2007/12	1	0	Creation of TAF
2008/01	1	0	Bank of America acquisition of Countrywide
2008/02	0	0	Northern Rock taken into state ownership
2008/03	0	1	Sale of Bear Sterns to JPMorgan
2008/04	1	1	FOMC rate cut
2008/05	0	1	FOMC expands the list of eligible collateral for TSLF
2008/06	1	0	S&P downgrade of AMBAC and MBIA
2008/07	1	1	Failure of IndyMac
2008/08	1	1	FOMC maintains federal funds rate target
2008/09	1	1	Lehman Bankruptcy
2008/10	0	0	Establishment of TARP
2008/11	1	0	Creation of TALF
2008/12	1	1	Fed extension of liquidity facility
2009/01	0	1	Treasury, Fed, FDIC assistance to BofA
2009/02	1	1	Increase of TALF to $1 trillion
2009/03	0	1	
2009/04	0	0	
2009/05	1	1	
2009/06	0	0	Large banks' repayments of bailout funds
2009/12	1	0	Escalation of Greek debt crisis
2010/04	0	0	Greece seeks financial support
2010/05	1	0	Establishment of EFSM and EFSF
2010/09	1	1	Second Greek bailout installment
2011/06	1	0	Merkel agrees to Greece bondholder role
2011/10	1	0	Call for a Greek referendum