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

Hui Chen Scott Joslin Sophie Ni*

November 14, 2014

Abstract

We propose a new measure for the variations in intermediary constraints by observing how financial intermediaries manage their tail risk exposures. Using a unique dataset on the trading activities between public investors and financial intermediaries in the market for deep out-of-the-money S&P 500 put options, we are able to isolate shocks to intermediary constraints by exploiting the price-quantity relations in this market. Besides the effects on option pricing, our measure of the shocks to intermediary constraints is a strong predictor of future returns for a wide range of financial assets, and it is associated with changes in the aggregate intermediary leverage. To explain these findings, we build a general equilibrium model of the crash insurance market, where time variation in intermediaries' constraints help generate the dynamic relationships between equilibrium public demand for crash insurance, intermediary leverage, and the aggregate risk premium.

*Chen: MIT Sloan and NBER (huichen@mit.edu). Joslin: USC Marshall (sjoslin@usc.edu). Ni: Hong Kong University of Science and Technology (sophieni@ust.hk). We thank Tobias Adrian, David Bates, Robert Battalio, Geert Bekaert, Harjoat Bhamra, Nicolae Garleanu, Zhiguo He, Bryan Kelly, Lei Lian, Andrew Lo, Dmitriy Muravyev, Jun Pan, Lasse Pedersen, Steve Ross, Ken Singleton, Moto Yogo, Hao Zhou, and seminar participants at USC, MIT Sloan, HKUST, Notre Dame, Washington University in St. Louis, New York Fed, Boston Fed, Copenhagen Business School, AFA, the Consortium of Systemic Risk Analytics Meeting, CICF, SITE, the OptionMetrics Research Conference, and the Risk Management Conference at Mont Tremblant for comments, and Ernest Liu for excellent research assistance. We also thank Paul Stephens at CBOE, Gary Katz and Boris Ilyevsky at ISE for providing information on the SPX and SPY option 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 by observing how financial intermediaries manage their tail risk exposures. Using a unique dataset on the trading activities between public investors and financial intermediaries in the market for deep out-of-the-money put options on the S&P 500 index (henceforth referred to as DOTM SPX puts, which are effectively insurance against large market crashes), we are able to isolate periods with the presence of supply shocks, i.e., shocks to the intermediaries' capacity to provide crash insurance. We then show that the tightening of intermediary constraint is associated with increasing expensiveness of DOTM puts, rising risk premia for a wide range of financial assets, deterioration in funding liquidity, as well as de-leveraging by broker-dealers. Finally, we present a general equilibrium model with time-varying intermediary constraints that captures the main empirical findings.

Our main measure is the net amount of DOTM SPX puts that public investors in aggregate acquire 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 turn negative during times of market distress. Our identification strategy exploits the basic relation between the quantities of trading, measured by PNBO, and prices, based on two measures of option expensiveness. While a positive price-quantity relation is consistent with the presence of shocks to public investors' demand for crash insurance, a negative price-quantity relation is consistent with the presence of shocks to financial intermediaries' capacity to provide crash insurance.[1]

In monthly data from January 1991 to December 2012, PNBO is significantly negatively

[1] We assume that public investors' demand curve is downward sloping, and financial intermediaries' supply curve is upward sloping. These assumptions are true when public investors and financial intermediaries are (effectively) risk-averse and cannot fully unload the inventory risks through dynamic hedging. Notice that supply shocks and demand shocks are not mutually exclusive. For example, a negative price-quantity relation does not rule out the presence of demand shocks.

related to the 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 options market. They are consistent with shocks to intermediary constraints being a main driving force behind changes in option premium and quantity of trading between public investors and financial intermediaries in the majority of our sample period.

Next, we show that PNBO significantly predicts future market excess returns. In the full sample, a one-standard deviation decrease in PNBO is on average associated with a 3.4% increase in the subsequent 3-month market excess return. The R^2 of the return-forecasting regression is 17.4%. The predictive power of PNBO is even stronger in the months when the PNBO/option-expensiveness relation is negative, where the R^2 is 22.8%, and it is weaker in the months when the PNBO/option-expensiveness relation turns positive. Besides equity, a lower PNBO predicts higher future returns on high-yield corporate bonds, commodity, carry trade, and an aggregate hedge fund portfolio, and it predicts low returns on long-term Treasuries.

The return predictability results above are consistent with the intermediary asset pricing theories, where the reduced risk-sharing capacity of the intermediaries raises the aggregate risk premium in the economy. Moreover, our results suggest that it is the combination of a negative price-quantity relation and a lower public net-buying volume in the market for crash insurance that best identifies the tightening of intermediary constraints.

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 as well as intermediary constraints. This is in contrast to the variation in intermediary constraints having a direct impact on the aggregate risk premium, which is a crucial distinction for intermediary asset pricing theories. If this alternative explanation is true, then the inclusion of the proper risk factors into the predictability regression should drive away the predictive power of PNBO. We find that the predictive power of PNBO is unaffected by the inclusion of a long list of return predictors in the literature, including various price ratios, the consumption-wealth ratio, the variance risk premium, the default spread, the term spread, and a variety of 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 the intermediary constraints having an unique effect on the aggregate risk premium.

Our PNBO measure is related to a list of funding liquidity measures in the literature. Specifically, PNBO is negatively correlated with the CBOE VIX index, the funding liquidity measure of Fontaine and Garcia (2012), the illiquidity measure of Hu, Pan, and Wang (2013), and it is significantly positively correlated with the growth rate in broker-dealer leverage, a funding constraint measure advocated by Adrian and Shin (2010) and Adrian, Moench, and Shin (2010). When regressing market excess returns jointly on lagged PNBO and other funding constraint measures, the coefficient on PNBO remains significant, which suggests that PNBO contains unique information about the aggregate risk premium relative to the other funding liquidity measures.

Our analysis of the price-quantity dynamics suggests that when financial intermediaries switch from sellers of DOTM SPX puts to buyers (e.g., in the months following the Lehman Brothers bankruptcy), 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. As part of our understanding of the risk sharing mechanism, it is important to identify which types of public investors (retail or institutional) are selling the DOTM puts to the intermediaries during times of distress. While the CBOE data do not reveal the types of public investors directly, we answer this question by comparing the predictive power of PNBO measures constructed based on 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), as well as based on large vs. small public orders of SPX options (with the large orders more likely from institutional investors). The results from both tests suggest that it is the institutional investors who are selling the DOTM puts to the financial intermediaries during periods of distress.

Finally, we present a dynamic general equilibrium model featuring time-varying intermediary constraints that can capture the main empirical findings in our paper. The purpose of the model is two-fold. First, it helps generate 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 than those predictions derived from intuition. Second, building on many existing models that emphasize the micro foundation of the intermediary constraints, we propose a reduced-form approach that directly links the tightness of intermediary constraints to intermediaries' effective risk aversion. This approach enables us to study the quantitative effects of intermediary constraints in a more tractable way and to develop new testable predictions.

In the model, public investors' equilibrium demand for crash insurance depends on the level of tail risk in the economy, the wealth distribution between public investors and the intermediaries, and shocks to the intermediation capacity. As the probability of a market crash rises, all else equal, public investors' demand for crash insurance tends to rise. However, if the intermediaries' risk sharing capacity drops at the same time due to loss of wealth or increase in risk aversion, they will choose to reduce their leverage, and the equilibrium amount of trading can become smaller. Furthermore, because of reduced risk sharing, public investors now demand a higher premium for bearing crash risk. Our calibrated model shows that this mechanism generates significant variation in the aggregate risk premium.

Our model builds on and extends the work of Garleanu, Pedersen, and Poteshman (2009) (henceforth GPP) to incorporate 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. Like GPP, the limited intermediation capacity of the dealers is a key feature of our model, but we introduce shocks to the intermediary risk-sharing capacity and endogenize the public demand for options, option pricing, and aggregate market risk premium in general equilibrium. In our empirical analysis, we 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 identification 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 seminar 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 Krishnamurhty (2012), Adrian and Boyarchenko (2012), Brunnermeier and Sannikov (2013), among others. Our reduced-form approach to model the intermediary constraint helps us derive new predictions on the effects of time-varying intermediary constraints in a tractable way.

In contrast to the fast growing body of theoretical work, there is relatively little empirical work on measuring the intermediary constraints and studying their aggregate effects on asset prices. The notable exceptions include Adrian, Moench, and Shin (2010) and Adrian, Etula, and Muir (2012), 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 particular mechanism (the crash insurance market) through which intermediary constraints affect aggregate risk sharing and asset prices. By utilizing the price-quantity information, we can better isolate shocks to intermediaries' risk sharing capacity. Moreover, compared to intermediary leverage changes, our measure has the advantage of being forward-looking and available at higher (daily vs. 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, our paper is related to several studies that 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 aggregate market risk premium is highly sensitive to the amount of sharing of tail risks in equilibrium.

2 Empirical Evidence

In this section, we present the empirical evidence connecting the trading activities of S&P 500 index options between public investors and financial intermediaries to the constraints of the financial intermediaries, the pricing of index options, and the risk premium of the aggregate stock market.

2.1 Data and variables

The data used to construct our measures of option trading activities are from the Chicago Board Options Exchange (CBOE). The Options Clearing Corporation 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. The option volume data are available daily from 1991 to 2012. Option pricing and open interest data are obtained from OptionMetrics from 1996 to 2012.

Since we want to link the option market trading activities to the constraints of financial intermediaries, it is natural to merge firm investors and market-makers as one group and observe how they trade against public investors. Our main option volume variable is the public net buying-to-open volume, or PNBO, which is defined as the total open-buy orders of all the deep out-of-the-money (DOTM) SPX puts by public investors minus their open-sell orders on the same set of options in each month. Since options are in zero net supply, the amount of net buying by the public investors is equal to the amount of net selling by firm investors and market-makers.

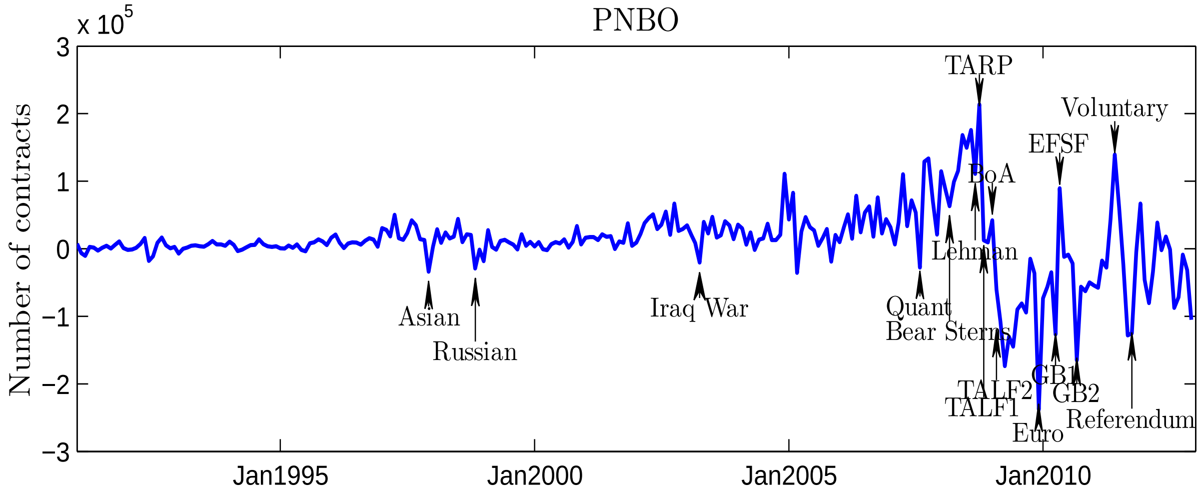
DOTM puts are defined as those with strike-to-price ratio $K/S \leq 0.85$. For robustness, later on we also present the results based on different strike-to-price cutoffs, as well as cutoffs that adjust for the option maturity and the volatility of the S&P 500 index. Our measure focuses on open orders (orders to open new positions) because they are not mechanically influenced by the existing orders (as in the case of close orders).[2] For comparison, we will also consider the following variations of the PNBO measure: (i) PNB, which is the public net buying volume (including both open and close orders); (ii) FNBO, which is the firm investors net buying-to-open volume; and (iii) PNBAll, which is the public net buying volume of all SPX puts (not just DOTM). Finally, we also define a normalized PNBO measure (PNBON), which is PNBO divided by the average of previous 3-month total public option volume.

Figure 1 plots the time series of PNBO and its normalized value PNBON. Consistent with the finding of Bollen and Whaley (2004) and GPP (who consider option demand across all strikes), 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 the broker-dealers and market-makers were mainly providing market crash insurance to the 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).[3]

However, starting in 2007, PNBO became significantly more volatile. It turned negative during the quant crisis in August 2007, when a host of quant-driven hedge funds experienced

[2]An investor might close an existing position because of its past performance or because it is approaching expiration, rather than based on new information or new hedging needs.

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



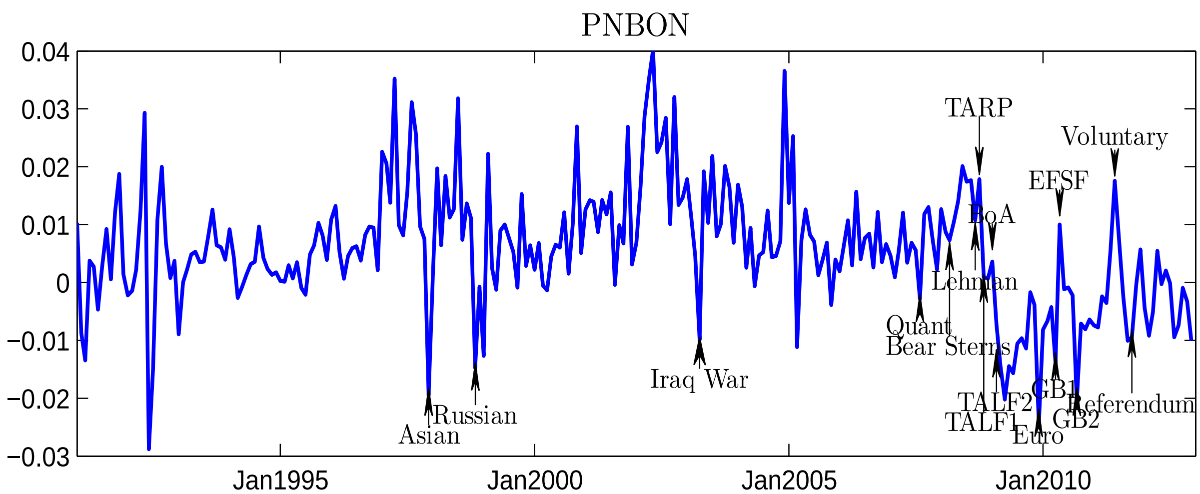


Figure 1: **Time Series of net public purchase for DOTM SPX puts.** PNBO is the net amount of DOTM (with $K/S \leq 0.85$) SPX puts public investors buying-to-open each month. PNBON is PNBO normalized by average of previous 3-month total volume from public investors. "Asian" (1997/12): period around the 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.

significant losses. It then rose significantly and peaked in October 2008, following the bankruptcy of Lehman Brothers. As financial market conditions continued to deteriorate, PNBO plunged rapidly and turned significantly negative in the following months. Following a series of government actions, PNBO first 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 a net amount of 44,000 DOTM SPX puts to open 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 over time. This consideration motivates us to normalize the PNBO series by the past 3-month average trading volume by public investors, yielding the PNBON series. As the bottom panel of Figure 1 shows, the PNBON series no longer demonstrates visible trend in volatility over time.

Our identification strategy focuses on the relation between PNBO and the expensiveness of SPX options. We use two measures of the expensiveness of SPX options. The first measure is the variance premium (VP) in Bekaert and Hoerova (2014), which captures the overall expensiveness of options at different moneyness. In each month, VP is the average of daily difference between VIX^2 and the expected physical variance based on a forecasting model with the best out-of-sample performance among a large group of models. Our second measure for option expensiveness is *Slope*, the monthly average of daily slope of the implied volatility curve for options expiring in the following month, which measures the expensiveness of DOTM puts relative to ATM options. Finally, we follow Andersen, Bollerslev, and Diebold (2007) to compute a monthly jump risk measure J, which is the average of daily physical jumps in the S&P 500 index.

Table 1 reports the summary statistics of the option volume and pricing variables. From January 1991 to December 2012, the net public open-purchase of the DOTM SPX puts (PNBO) is close to 10,000 contracts per month on average (each contract is 100 times the index). In comparison, the average total open interest for all DOTM SPX puts is

Table 1: **Summary Statistics**

This table reports the summary statistics for the main option volume and pricing variables in the empirical analysis. AC(1) is the first order autocorrelation; pp-test is the p-value for the Phillips-Perron test for unit root. PNBO: net open-buying volume of DOTM index puts ($K/S <= 0.85$) by public investors. PNBON: PNBO normalized by past 3-month average total options volume. PNB: public net open- and close-buying volume. FNBO: net open-buying volume by firms. PNBAll: public net buying volume of all SPX puts. OpenInt: end-of-month total open interest for all DOTM SPX puts. VP: variance premium from Bekaert and Hoerova (2014). *Slope*: difference in the implied volatility between one-month DOTM and ATM SPX puts. J is a measure for physical jump risk from Andersen, Bollerslev, and Diebold (2007). OpenInt is from 1996/01-2012/12. The other data series are for the period 1991/01-2012/12.

	mean	median	std	AC(1)	pp-test
PNBO (contracts)	9996	9665	51117	0.61	0.00
PNBON (%)	0.414	0.396	0.747	0.45	0.00
PNB (contracts)	25799	12599.5	55634	0.37	0.00
FNBO (contracts)	2686	-42.5	40567	0.37	0.00
PNBAll (contracts)	83623	57736	95306	0.41	0.00
OpenInt (contracts)	919923	507969	1030913	0.67	0.00
VP	13.02	10.39	20.21	0.33	0.00
Slope	29.67	29.50	6.03	0.56	0.21
J	4.50	3.75	5.07	0.33	0.00

around 0.9 million contracts during the period from January 1996 to December 2012. For the whole sample, the correlation between PNBO and FNBO (net open-purchase by firm investors) is -0.4, suggesting that firm investors tend to trade against the public investors as a whole. The option volume measures have relatively modest autocorrelations (0.61 for PNBO and 0.37 for PNB) compared to standard return predictors such as dividend yield and term spread.

The PNBO series is pro-cyclical, as indicated by its positive correlation with industrial production growth (0.17) and negative correlation with the unemployment rate (-0.48). Adrian and Shin (2010), Adrian, Moench, and Shin (2010), and Adrian, Etula, and Muir (2012) argue that the low (negative) balance sheet growth of the financial intermediaries is a sign of tight financing constraints. The correlation between PNBO and the year-over-year growth rate in broker-dealer leverage (Δlev) is 0.5, which implies that during times when

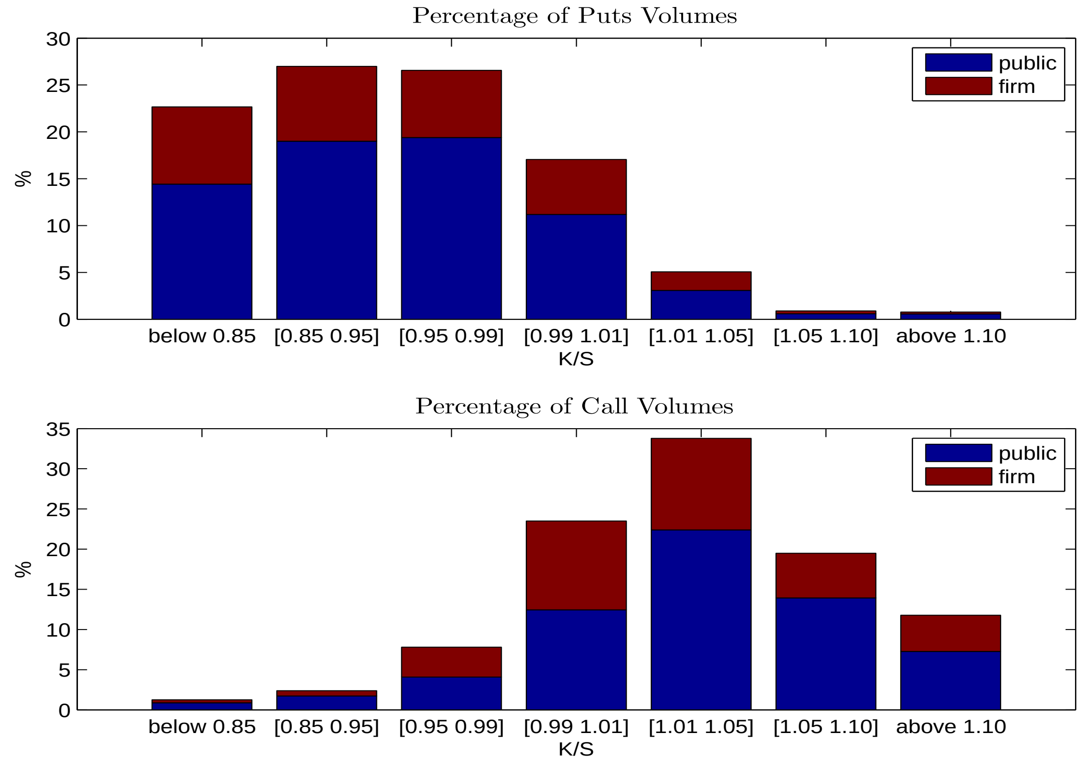


Figure 2: **Percentage of total put and call volumes at different moneyness.**

broker-dealers and market-makers sell a smaller amount of DOTM SPX puts or even become net buyers of these options, they tend to experience negative changes in leverage.

Figure 2 provides information about the trading volume of SPX options at different moneyness. Over the entire sample, put options account for 63% of the total trading volume of SPX options by public and firm investors. Within 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.

Besides being a large and active market, the market for DOTM SPX puts is well suited for our empirical investigation for two additional reasons. First, 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 puts, which are highly sensitive to jump risk which is difficult to hedge. Second, while SPX puts

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

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. The sample period is $1996-2012$.

	$\frac{K}{S}<0.85$		$0.85<\frac{K}{S}<0.95$		$0.95<\frac{K}{S}<0.99$		$0.99<\frac{K}{S}<1.01$	
	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

are not the only financial instruments to hedge against tail risks, they provide unique advantages over OTC derivatives (e.g., credit derivatives) in that the central counterparty clearing and margin system largely remove the counterparty risks and enhance liquidity, which are particularly relevant when dealing with aggregate tail risk.

To demonstrate the first point above, we regress put option returns on the returns of the corresponding hedging portfolios at both weekly and daily horizons, and we restrict the options to be between 15 and 90 days to maturity to ensure liquidity. We consider two hedging portfolios, one based on delta hedging (using the underlying S&P500 index) and one based on delta-gamma hedging.[4] The R^2s of these regressions demonstrate how effective the two types of hedging methods are.

As Table 2 shows, at daily (weekly) horizon, 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 the R^2 is lower than 60% for DOTM puts, and no more than 75% for OTM puts with K/S between 0.85 and 0.95. These results imply that when holding non-zero inventories of OTM SPX puts, especially DOTM puts, the financial intermediaries will be exposed to significant inventory risks even after dynamically hedging these positions. We are interested in studying how the inventory risks combined with the time-varying constraints facing the financial intermediaries affect option pricing and the risk sharing in the economy.

[4]For delta-gamma hedging, we use at-the-money puts expiring in the following month in addition to the S&P500 index.

2.2 Option volume and the expensiveness of SPX options

We start by investigating the link between PNBO and the expensiveness of SPX options. As explained earlier, we use two different measures for the expensiveness of SPX options. First, we use the variance premium (VP) from Bekaert and Hoerova (2014), which measures the overall expensiveness of SPX options. This measure compares the average risk-neutral variance (as measured by the VIX) with the expected physical variance.[5] This measure thus capture the expensiveness of the options by comparing prices (through VIX) to a forward looking measure of the risk exposure in the options. Additionally, this measure of the variance premium using forward-looking measure of physical variance is related to future stock market returns, which we will analyze in our subsequent analysis. The second measure of expensiveness that we use is the average daily slope of the implied volatility curve ($Slope$), which measures the relative expensiveness of one-month DOTM puts compared to ATM puts.

We use the following regressions to examine the relations between quantities and prices:

$$VP_t = a_{VP} + b_{VP}\ PNBO_t + c_{VP}\ PNBO_t \times J_t + \epsilon_t^v, \text{ and} \tag{1}$$

$$Slope_t = a_{Slope} + b_{Slope}\ PNBO_t + c_{Slope}\ PNBO_t \times J_t + \epsilon_t^s. \tag{2}$$

Additionally, we suppose that the relationship between equilibrium demand and option expensiveness may depend on the amount of jump risk. Thus, in both (1) and (2), we interact PNBO with a jump risk measure J from Andersen, Bollerslev, and Diebold (2007).

The sign of b_{VP} in (1) and b_{Slope} in (2) can help us distinguish between demand shocks and supply shocks in the options market. The demand pressure theory of GPP predicts that a positive and exogenous shock to the public demand for DOTM put options forces risk-averse dealers to bear more inventory risks. As a result, the dealers will raise the price of the option (a move on the upward-sloping supply curve). Since the unhedgeable parts

[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.

of the DOTM puts are much more significant than those of the ATM puts, the DOTM puts should also become more expensive relative to the ATM puts. Thus, demand shocks will lead to $b_{VP} > 0$ and $b_{Slope} > 0$. Furthermore, the effect of public demand on option prices should be stronger when jump risks in the market are high, because those are times when dealers are particularly concerned with inventory risks (the supply curve becomes steeper), especially in the case of DOTM puts. This implies $c_{VP} > 0$ and $c_{Slope} > 0$.

Alternatively, if there are intermediation shocks that raise the degree of constraints facing financial intermediaries (e.g., due to loss of capital or tightened 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 public demand falls endogenously, implying that $b_{VP} < 0$ (a move on the downward-sloping demand curve). Moreover, the fact that the unhedgeable inventory risks apply more to DOTM puts than ATM puts implies that $b_{Slope} < 0$. The sign of c_{VP} depends on how the slope of the demand curve changes with higher jump risks. Since the demand for crash insurance is likely more sensitive to a rise in jump risk at times when the level of crash risk is high, the demand curve will become steeper, which implies $c_{VP} < 0$.[6]

Table 3 reports the results based on monthly data. The top panel uses variance premium (VP) to measure the expensiveness of SPX options. Whether we use PNBO or the normalized PNBO (PNBON) as the measure of public net buying volume for DOTM SPX puts, the coefficient b_{VP} is negative and statistically significant, consistent with the hypothesis that changes in the equilibrium quantities of DOTM SPX puts that public investors purchase are mainly driven by shocks to intermediaries' capacity to provide crash insurance. In the univariate regression, a one standard deviation increase in PNBO is associated with a one-third standard deviation decrease in variance premium. Adding the interaction term between $PNBO_t$ and the jump risk measure, J_t, significantly raises the adjusted R^2 relative to the univariate regression. The coefficient c_{VP} on the interaction term is significantly negative, which is also consistent with the intermediary constraint

[6]Our reasonings for the signs of the regression coefficients are admittedly informal. Besides, the sign of c_{Slope} in the case of supply shocks is difficult to determine based on intuition. We formally investigate these properties in the dynamic model in Section 3.

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

The dependent variables are (i) VP: the variance premium in Bekaert and Hoerova (2014), and (ii) *Slope*: the implied volatility of one-month DOTM SPX puts ($K/S < 0.85$) minus that of ATM puts. We use three different measures of the public net-buying volumes: PNBO (public net buying-to-open volume for DOTM puts), PNBON (PNBO normalized by past 3-month average public volume), and PNBAll (public net buying volume of all SPX puts). J is the average of daily physical jump of S&P 500 in Andersen, Bollerslev, and Diebold (2007). Standard errors in parentheses have corrected for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). We use monthly data from January 1991 to December 2012.

	PNBO		PNBON		PNBAll	
	$VP_t = a_{VP} + b_{VP}\ PNBO_t + c_{VP}\ PNBO_t \times J_t + \epsilon_t^v$					
b_{VP}	-117.26	-84.06	-4.49	-3.76	-40.79	-47.89
	(40.48)	(19.38)	(1.35)	(0.95)	(16.79)	(41.02)
c_{VP}		-6.45		-0.32		-1.46
		(2.16)		(0.12)		(5.56)
Adj. R^2	0.09	0.17	0.09	0.16	0.02	0.03
	$Slope_t = a_{Slope} + b_{Slope}\ PNBO_t + c_{Slope}\ PNBO_t \times J_t + \epsilon_t^s$					
b_{Slope}	-24.63	-25.98	-1.19	-1.19	2.29	-1.63
	(5.35)	(5.05)	(0.29)	(0.28)	(3.25)	(3.77)
c_{Slope}		0.26		0.00		-0.80
		(0.24)		(0.02)		(0.28)
Adj. R^2	0.11	0.11	0.12	0.12	0.00	0.06

theory, in particular with the hypothesis that demand curve steepens when jump risk rises.

Next, when we measure the option expensiveness using the difference in the implied volatility between DOTM and ATM SPX puts (*Slope*), the results are again consistent with the intermediary constraint theory. In particular, b_{Slope} is significantly negative, which is consistent with intermediation shocks having a larger impact on the pricing of DOTM puts than on ATM puts since they are more difficult to hedge (see Table 2).

When we use public net buying volume for all SPX options (PNBAll) to measure public demand instead of PNBO, not only are the R^2 of the regressions much smaller, but the regression coefficients are no longer significantly different from zero in several cases (including b_{VP} and c_{VP} in regression (1), and b_{Slope} in regression (2)). The fact that

PNBO is more strongly connected to option expensiveness than PNBOAll is consistent with the interpretation that trading activities in the market for DOTM SPX puts contain unique information about financial intermediary constraints relative to the markets for other options (again due to the fact that DOTM puts are more difficult to hedge).

Our empirical specifications in (1) and (2) are connected to that of Garleanu, Pedersen, and Poteshman (2009). Consistent with the demand pressure theory, GPP find a positive relation between their measure of option expensiveness and public net open interest. Our regressions differ from GPP in several aspects. First, our PNBO measure uses contemporaneous public net-buying volume, while GPP use public net open interest, which is the accumulation of past net-buying volumes. At monthly frequency, this difference between net-buying volume and net open interest does not apply for options with maturities of one month or less. Yet we find similar results as in Table 3 when we compute PNBO using only these short-dated options. Second, our PNBO measure focuses on DOTM puts, whereas GPP use options of all moneyness (similar to our PNBOAll in this regard). Third, our sample period is 1991-2012, while theirs is 1996-2001.

The results in Table 3 are not a rejection of the effect of demand shocks on option prices. 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. In fact, according to our identification strategy, supply shocks and demand shocks are not mutually exclusive. A negative price-quantity relation does not rule out the presence of demand shocks. It only implies that the effect of supply shocks is dominant. Similarly, a positive price-quantity relation does not rule out the presence of supply shocks.

Furthermore, in sub-periods when intermediary constraints are not varying strongly over time, PNBO and option premium can be uncorrelated or even positively correlated (due to demand shocks). To examine this point, we run regression (1) using daily data for each calendar month. Out of 264 months, the relation between VP and PNBO is negative in 159 months (b_{VP} is significant at 5% level in 44 of them) and positive in the remaining 105 months (b_{VP} is significant at 5% level in 24 of them). These results suggest that shocks to intermediary constraints are the main driving force behind the price-quantity dynamics in the majority of our sample period; moreover, these shocks are likely to be stronger during the periods with negative b_{VP}. We exploit this time variation in the price-quantity relation in the following analysis of PNBO and the aggregate risk premium.

2.3 Option volume and market risk premium

According to the theory of financial intermediary constraints (see e.g., Gromb and Vayanos (2002) and He and Krishnamurhty (2012)), variations in the aggregate intermediary constraints not only affect option prices, but also drive the risk premium of other financial assets. In this section, we examine the ability of PNBO in forecasting future excess returns on the CRSP value-weighted market portfolio. The basic specification of the return forecasting regression is:

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

where the notation $t+j \to t+k$ indicates the leading period from $t+j$ to $t+k$ ($k > j \geq 0$).

Table 4 shows PNBO has strong predictive power for future market excess returns up to 4 months ahead. The coefficient estimates are all negative and statistically significant when the dependent variables are the 1st, 2nd, 3rd, and 4th month market excess returns. The coefficient estimate for predicting one-month ahead market excess returns is -24.06 (t-stat $= -3.99$), with an R^2 of 7.8%. For 4-month ahead returns ($r_{t+3\to t+4}$, or simply r_{t+4}), the coefficient estimate is -17.26 (t-stat $= -2.10$), with an R^2 of 4.0%. From 5 months out to 10 months, the predictive coefficient b_r is still negative, but no longer statistically significant. When we aggregate the effect for the cumulative market excess returns in the next 3 months, the coefficient of -65.32 implies that a one-standard deviation decrease in PNBO raises the future 3-month market excess return by 3.4%. The R^2 is 17.4%.

Since Figure 1 indicates that non-stationarity might be a potential concern for PNBO, we also used the normalized PNBO to predict market excess returns. Table 4 shows that, like PNBO, PNBON also predicts future market returns negatively, and the coefficient b_r remains statistically significant at the 1% level up to 3 months ahead (based on the

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 market excess return from $t+j$ to $t+k$ ($k > j \geq 0$). Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). The calculation of bootstrap confidence intervals follows Welch and Goyal (2008). Sample period: 1991/01 - 2012/12.

	Return	b_r	$\sigma(b_r)$	99% bootstrap CI	R^2 (%)
	PNBO				
Full Sample	$r_{t\to t+1}$	-24.06	(6.03)	[-37.07, -10.67]	7.8
	$r_{t+1\to t+2}$	-18.55	(5.17)	[-31.64, -5.09]	4.7
	$r_{t+2\to t+3}$	-23.32	(5.78)	[-36.59, -10.02]	7.3
	$r_{t+3\to t+4}$	-17.26	(8.23)	[-30.55, -3.82]	4.0
	$r_{t\to t+3}$	-65.32	(15.72)	[-88.19, -42.35]	17.4
Sub-sample: $b_{VP} < 0$	$r_{t\to t+3}$	-81.12	(17.25)	[-128.55, -44.57]	22.8
Sub-sample: $b_{VP} > 0$	$r_{t\to t+3}$	-43.85	(13.84)	[-71.83, 0.94]	9.6
	PNBON				
Full Sample	$r_{t\to t+1}$	-0.90	(0.30)	[-1.65, -0.17]	4.6
	$r_{t+1\to t+2}$	-0.65	(0.26)	[-1.31, -0.01]	2.4
	$r_{t+2\to t+3}$	-0.72	(0.28)	[-1.45, -0.06]	3.0
	$r_{t+3\to t+4}$	-0.57	(0.27)	[-1.24, 0.10]	1.9
	$r_{t\to t+3}$	-2.27	(0.78)	[-3.46, -1.07]	8.9
Sub-sample: $b_{VP} < 0$	$r_{t\to t+3}$	-2.60	(0.52)	[-3.92, -1.36]	11.8
Sub-sample: $b_{VP} > 0$	$r_{t\to t+3}$	-1.56	(1.06)	[-4.03, 1.01]	4.1

bootstrapped confidence interval).

In Section 2.2 we show that while the overall relation between measures of option expensiveness and PNBO is negative, this relation varies in sub-periods. To the extent that periods with a positive price-quantity relation are associated with weaker (or a lack of) shocks to intermediary constraints, the predictive power of PNBO should be weaker (no longer exist) during such times.

The sub-sample analysis in Table 4 verifies this prediction. For both PNBO and PNBON, the regression coefficient b_r and the R^2 are higher in the sub-sample periods

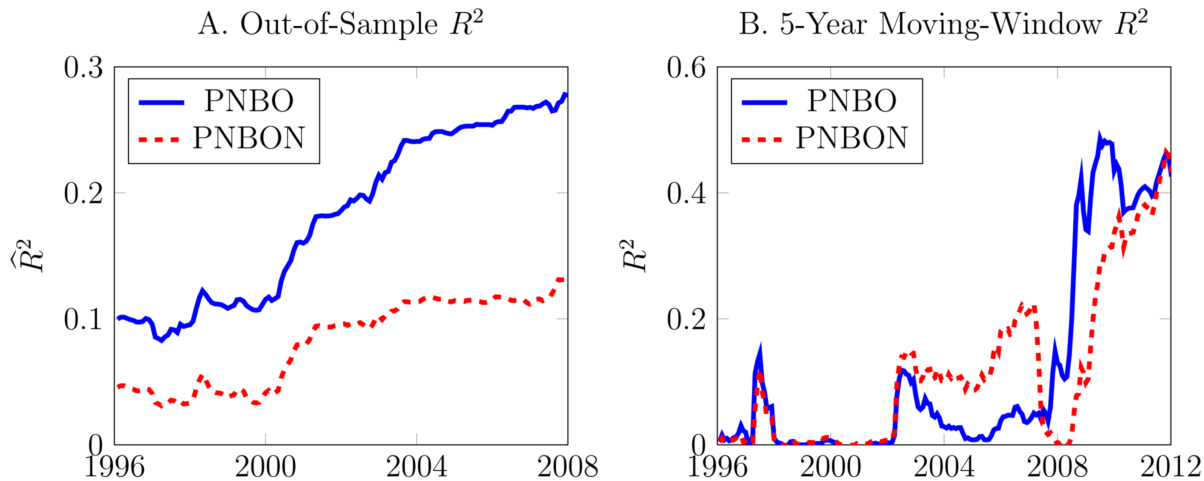


Figure 3: **Out-of-sample R^2 and R^2 from 5-year moving-window regressions.** Panel A plots the out-of-sample R^2 as a function of the sample split date. Panel B plots the in-sample R^2 of 5-year moving-window regressions. In both panels, the return predictors are PNBO and PNBON, and the returns are 3-month market excess returns.

where $b_{VP} < 0$ than in the full sample. In contrast, the R^2 are lower in the sub-sample with $b_{VP} > 0$, and the bootstrapped 99% confidence intervals for the coefficient b_r are no longer strictly negative. These results show that it is the combination of a negative price-quantity relation and a smaller PNBO that best identifies the tightening of intermediary constraints.

To further investigate the reliability of the predictability results, 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). This is because recent studies suggest that sample splits themselves can be data-mined (Hansen and Timmermann (2012)). We first estimate the predictability regression for 3-month market excess returns during the estimation period, and then compute the mean squared forecast errors for the predictability model (MSE_A) and the historical mean model (MSE_N) in the evaluation period. Then, the out-of-sample R^2 is

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

Panel A of Figure 3 shows the results. PNBO achieves an out-of-sample R^2 above 10% for most sample splits, which remains above 20% from 2004 onward. Its normalized value PNBON has an an out-of-sample R^2 above 5% for most sample splits, which remains above 10% in the later period.

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 R^2 varies significantly over time. It is generally lower in the early parts of the sample, being less than 5% most of the time prior to 2006. It rises to 15% in the period around the Asian financial crisis and Russian default in 1997-98, then to above 10% around 2002 when the tech bubble burst. During the crisis period, the R^2 rises to close to 50%. Such high R^2 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 50% implies a Sharpe ratio for the market timing strategy that exceeds 1.2. The fact that such high Sharpe ratios persist during the financial crisis is again consistent with the presence of severe financial constraints that prevent arbitrageurs from taking advantages of these investment opportunities.

Next, we re-estimate the predictive regressions for 3-month market excess returns after deleting the most extreme observations of PNBO and PNBON (in terms of the absolute value). This test helps us assess to what extent the predictability results are driven by a small number of outliers, in particular those observations during the 2008-09 financial crisis. Figure 4 shows the 95% confidence interval of the b_r coefficient in predicting three-month future excess returns after the removal of the extreme observations of PNBO and PNBON. The figure shows that the regression coefficients have relatively stable point estimates and remain statistically significantly negative even after deleting the 40 most extreme observations (15% of the sample).

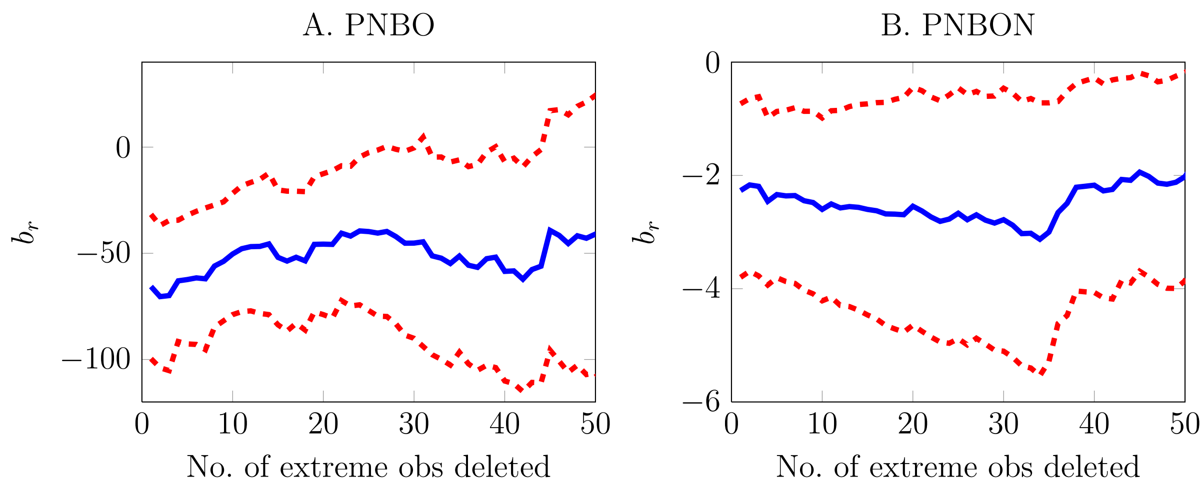


Figure 4: **Predictability of PNBO and PNBON after deleting extreme observations.** This figure plots the point estimates and 95% confidence intervals for b_r in the predictability regressions using PNBO and PNBON, after removing various numbers of extreme absolute PNBO observations. The dependent variable is the future three-month cumulative market excess returns.

Table 5 reports the results from several additional robustness tests. In the first row, we compare the predictive power of PNBO in the full sample (1991/01-2012/12) against the results from two sub-samples: pre-crisis (1991/01-2007/11) and post-crisis (2009/06-2012/12). The predictive power of PNBO remains statistically significant in both sub-samples but is weaker than the full sample (in terms of lower R^2 and smaller absolute value of b_r). These results show that while the relation between PNBO and market risk premium is stronger during the financial crisis, it is not a phenomenon that only occurs in the financial crisis. 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 as in the second half. Another possible reason is that the 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.

The PNBON measure is in part motivated by the concern of the growth of the option markets. It normalizes PNBO with the average of previous 3-month total public option volume. In Table 5, we consider a variation of PNBON, which is PNBO normalized by monthly average total SPX volume in the previous 12 months. The results are close to

Table 5: **Return Forecasts with Various SPX Option Volume Measures**

This table reports the results of the return forecasting regressions in different subsamples and on alternative option volume measures. r_{t+1} indicates market excess return one month ahead, whereas $r_{t\to t+3}$ indicates cumulative 3-month ahead market excess return. $PNBO/SPXVOL$ is PNBO normalized by the average total SPX volume in the previous 12 months. PNB is the public net buying volume (including both open and close orders). FNBO is the firm net buying-to-open volume. Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). Full sample period: 1991/01 - 2012/12. Pre-crisis: 1991/01 - 2007/11. Post-crisis: 2009/06 - 2012/12.

Return	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2
	$PNBO$ full-sample			$PNBO$ pre-crisis			$PNBO$ post-crisis		
$r_{t\to t+1}$	-24.06	(6.03)	0.08	-18.49	(9.17)	0.01	-15.82	(11.55)	0.04
$r_{t\to t+3}$	-65.32	(15.72)	0.17	-54.02	(19.62)	0.03	-40.48	(20.76)	0.13
	$PNBO/SPXVOL$			PNB			$FNBO$		
$r_{t\to t+1}$	-114.77	(39.04)	0.03	-22.72	(4.45)	0.08	13.74	(7.90)	0.01
$r_{t\to t+3}$	-291.16	(106.02)	0.07	-49.13	(16.63)	0.11	40.71	(18.35)	0.04

those for PNBO in Table 4.

PNBO reflects public investors' newly established positions. The net supply of DOTM puts from financial intermediaries in a given period not only includes the newly established positions, but also the changes in existing positions. To measure this supply, we compute PNB as the sum of net open- and close-buying volumes for public investors. Comparing PNB with PNBO, the coefficient b_r and R^2 are essentially the same for the one-month ahead return forecast. In the cumulative 3-month return forecast, the R^2 falls and b_r drops in absolute value. We also examine the predictability of the net-open-buying volume from firm investors (FNBO). We can see that the firm investors net demand predict returns positively. This result is consistent with the interpretation that both broker-dealers and market-makers have positions opposite to the public investors.

Finally, we examine how the predictive power of PNBO changes based on option moneyness. As discussed in Table 2, because DOTM puts are more difficult to hedge than ATM puts, they expose financial intermediaries to more inventory risks. Hence, the PNBO

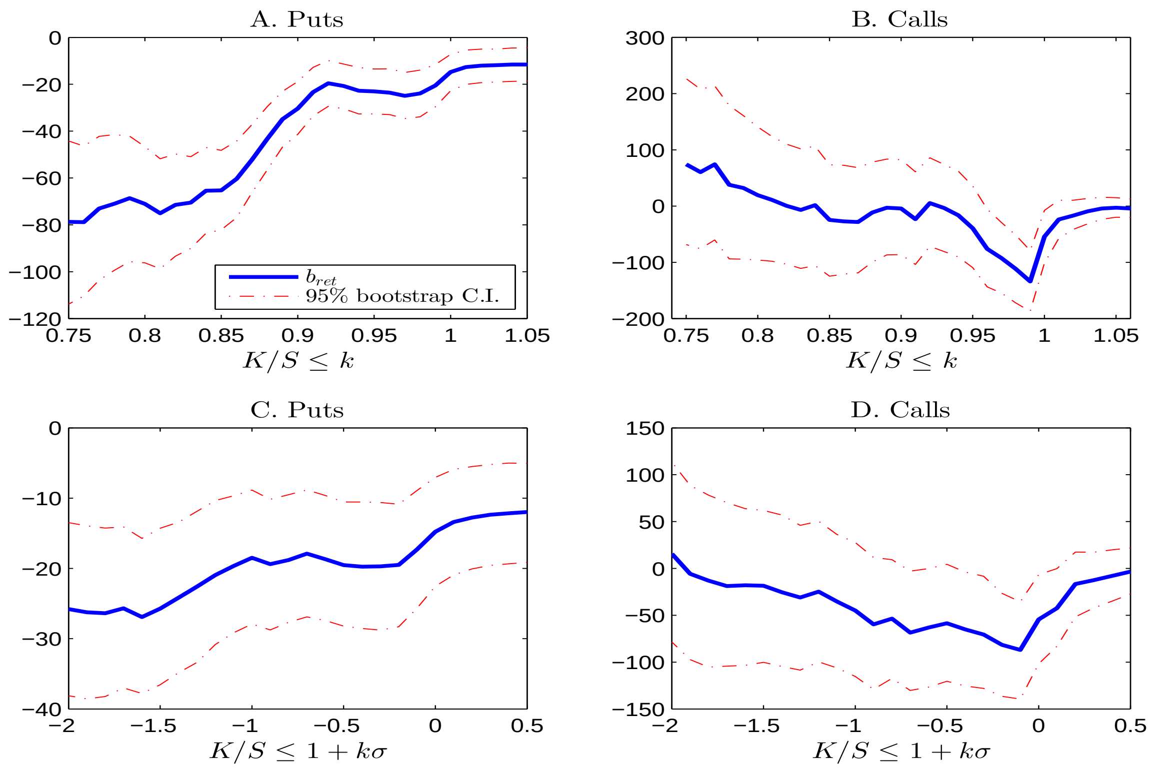


Figure 5: **Predictive power of PNBO at different moneyness.** In Panels A and B, PNBO is measured based on K/S less than a constant cutoff k. In Panels C and D, PNBO is measured based on K/S less than $1 + k\sigma$, where k is a constant, and σ represents a maturity-adjusted return volatility, which is the daily S&P return volatility in the previous 30 trading days multiplied by the square root of the days to maturity for the option.

measure based on DOTM puts should be more informative about intermediary constraints and the aggregate risk premium than those based on ATM puts. Our baseline definition of DOTM puts uses a simple cutoff rule $K/S \leq 0.85$. Panel A of Figure 5 plots the coefficient b_r and the confidence intervals as we vary the cutoff value. 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, which is consistent with the interpretation that the markets of DOTM puts are more sensitive to intermediary constraints. At the same time, because far out-of-money options are more thinly traded, the PNBO series becomes more noisy, which widens the confidence interval on b_r. In contrast, for almost all moneyness cutoffs, a PNBO measure based on SPX calls does not predict 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 some periods, but it might be more likely in other periods. For this reason, we 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. Panel C of Figure 5 shows that this alternative classification of DOTM puts produces qualitatively similar results as the simple cutoff rule. Again, Panel D shows that the PNBO series based on call options using this definition of the moneyness cutoff does not predict returns.

In summary, the fact that the predictive power of PNBO is the strongest for deep out-of-the-money puts is another piece of evidence that supports PNBO being a proxy for intermediary constraint.

2.4 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, e.g., He and Krishnamurhty (2012) and Adrian and Boyarchenko (2012).

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 variance premium (VP) in Bekaert and Hoerova (2014),[7] the log price-to-earning ratio ($p - e$) and 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 (2012), three measures of the slope of the implied volatility curve, 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 available quarterly.

Table 6 shows that, with the inclusion of the alternative predictive variables, the coefficient of PNBO remains significantly negative, and its size is largely unchanged across all the regressions. In contrast, only VP, $p - e$, IVSlope between $K/S \leq 0.85$ and $K/S \in (0.85, 0.95)$, and $\widehat{cay}$ are still statistically significantly related to future market returns in our sample period after PNBO is included in the regression. Comparing the R^2 from the multivariate regressions and the univariate regression, we see that the incremental explanatory power for future market excess returns mostly comes from PNBO.

Table 6 also shows that, unlike PNBO, the open interest for DOTM SPX puts has no predictive power for market excess returns. This result highlights that it is not the general trading activities for DOTM SPX puts, but rather the net exposures of the public investors and financial intermediaries that contain information about the market risk premium.

In summary, the results from Table 6 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.

[7] We have also used a related variance premium measure by Bollerslev, Tauchen, and Zhou (2009) (IVRV), which is the difference between implied volatility and historical volatility. The results based on VP and IVRV are similar.

Table 6: **Return Forecasts with PNBO and Other Predictors**

This table reports the results of forecasting 3-month market excess returns with PNBO and other predictors, including Variance premium (VP), log price-earnings ratio ($p-e$), log dividend yield ($d-p$), log net payout yield (lcrspnpy), Baa-Aaa credit spread (DEF), term spread (TERM), tail risk measure (Tail), three implied volatility slope measures (based on the differences in implied vol among 4 groups: $K/S \leq 0.85$, $K/S \in (0.85, 0.95)$, $K/S \in (0.95, 0.99)$, and $K/S \in (0.99, 1.01)$), and the quarterly consumption-wealth ratio ($\widehat{cay}$). Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). Sample period is 1991-2012, except for lcrspnpy and Tail (1991-2010), OpenInt (1996-2012), and the three IVSlope measures (1996-2012).

PNBO	-59.44	-70.13	-62.10	-69.59	-66.76	-68.69	-70.52	-62.29	-64.65	-63.96	-55.40	-87.29	
	(15.08)	(15.03)	(17.95)	(20.40)	(15.96)	(15.95)	(17.63)	(14.48)	(14.18)	(15.39)	(18.14)	(20.47)	
VP	0.05										0.06		
	(0.02)										(0.04)		
$p-e$		-3.42									-4.45		
		(1.92)									(2.29)		
$d-p$			2.76								4.79		
			(2.48)								(4.59)		
lcrspnpy				5.38							6.95		
				(2.15)							(1.94)		
DEF					-1.31						-5.13		
					(1.96)						(1.68)		
TERM						-0.66					-0.03		
						(0.73)					(0.58)		
Tail							31.58				-43.22		
							(22.23)				(34.26)		
IVSlope1								0.23			0.15		
								(0.10)			(0.10)		
IVSlope2									0.21		-0.60		
									(0.85)		(0.96)		
IVSlope3										1.97	1.41		
										(1.95)	(2.21)		
$\widehat{cay}$												45.32	
												(27.77)	
OpenInt													0.43
													(0.69)
R^2	0.19	0.18	0.18	0.21	0.17	0.18	0.18	0.19	0.18	0.19	0.32	0.17	0.00
Obs	264	264	264	240	264	264	240	204	204	204	180	87	204

2.5 Option volume and the returns of various assets

Having examined the ability of PNBO to predict future market excess returns, we now broaden the predictability regressions to other asset classes.

Among the asset classes we consider are 4 finance-related industry portfolios, 4 of the Fama-French 25 portfolios based on size and book-to-market, 2 momentum portfolios, a carry trade return series (computed based on the method in Bakshi and Panayotov (2013), i.e., the equally weighted log returns of four currency pairs from the G-10 currencies sorted on forward differentials),[8] a hedge fund portfolio (we use the HFRI fund-weighted average returns of all hedge funds), commodity (based on the Goldman Sachs commodity index excess return series), high-yield bonds (based on the Barclays U.S. Corporate High Yield total return index), and the 10-year US Treasury. Data for the returns on high yield bonds, commodity, and hedge funds are from Datastream. Government bond return data are from Global Financial Data. Returns on industry portfolios, momentum portfolios, and size/book-to-market portfolios are obtained from Ken French's website.

As Table 7 shows, PNBO predicts the future returns of a wide range of assets. Among the stock portfolios considered, the R^2 is higher for the 4 finance-related industry portfolios and the low-momentum portfolio (recent loser portfolios). Outside of equity, PNBO significantly predicts the returns on carry trade returns (3-month), hedge fund returns (1-month and 3-month), commodity returns (3-month), and returns on high yield corporate bonds (1-month and 3-month). The sign of the coefficient on PNBO is negative for all of these assets. Thus, like the market portfolio, the risk premia on these assets rise when the net amount of the crash insurance public investors buy from financial intermediaries become smaller.

The one exception is the 10-year Treasury. When predicting future returns on the 10-year Treasury, the coefficient b_r is positive and statistically significant at both 1-month and 3-month horizon. This result is consistent with the well-known "flight-to-quality" phenomenon. As the intermediary constraint tightens, aggregate market premium rises, and the demand for safe assets such as Treasuries also rises. As a result, the expected

[8] We thank George Panayotov for sharing the data.

Table 7: **Predicting Other Asset Returns with PNBO**

This table reports the results of forecasting future excess returns on a variety of assets and future VIX changes using PNBO. PNBO is public net buying-to-open volume for DOTM SPX puts. $r_{t+j\to t+k}$ represents excess return from $t+j$ to $t+k$ ($k > j \geq 0$). Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). Sample period: 1991/01 - 2012/12.

Assets	b_r	$\sigma(b_r)$	R^2 (%)	b_r	$\sigma(b_r)$	R^2 (%)
		$r_{t\to t+1}$			$r_{t\to t+3}$	
Industry: Banks	-34.3	(8.0)	7.8	-91.5	(26.1)	17.2
Industry: Insurance	-33.1	(8.3)	9.7	-85.0	(23.4)	20.4
Industry: Real Estate	-49.5	(17.0)	11.0	-135.0	(37.5)	22.3
Industry: Finance	-34.2	(14.0)	5.9	-89.5	(38.6)	11.9
Small Growth	-32.2	(6.7)	3.9	-88.6	(21.0)	8.1
Small Value	-32.1	(9.3)	7.7	-97.8	(25.5)	17.4
Large Growth	-21.2	(6.8)	5.6	-55.6	(19.5)	12.2
Large Value	-25.7	(7.5)	5.8	-68.3	(22.1)	12.1
Low Momentum	-58.8	(21.0)	9.9	-130.0	(54.3)	13.8
High Momentum	-21.5	(7.2)	3.0	-70.6	(19.2)	9.5
Carry trade	-6.7	(5.0)	1.4	-21.2	(13.4)	4.0
HFR hedge fund index	-8.9	(4.6)	5.0	-27.1	(14.1)	10.4
Goldman commodity index	-16.6	(15.0)	1.8	-62.7	(38.4)	6.7
Barclays corporate high yield index	-19.2	(7.8)	14.7	-48.9	(16.3)	22.0
10-year Treasury	7.1	(3.1)	2.8	16.3	(4.8)	4.9
		$\Delta_{t,t+1}$			$\Delta_{t,t+3}$	
VIX	18.1	(6.8)	4.8%	42.73	(13.9)	10.1%

return for holding Treasuries becomes lower during such times. Finally, PNBO also predicts future changes in VIX positively. This result is again consistent with low (negative) PNBO being an indicator for tighter intermediary constraint now. As the intermediary constraint relaxes over time, we expect the VIX index to fall.

2.6 Option volume and measures of funding constraints

Recently, several measures of funding constraints for financial intermediaries have been proposed in the literature. They include the balance sheet growth measures advocated by Adrian, Moench, and Shin (2010) and Adrian, Etula, and Muir (2012) (Δlev), the fixed-income market based funding liquidity measures by Fontaine and Garcia (2012) (FG) and Hu, Pan, and Wang (2013) (Noise), the CBOE VIX index (VIX), the TED spread (TED, the difference between the 3-month LIBOR and the 3-month T-bill rate), and the LIBOR-OIS spread (LIBOR-OIS, the difference between the 3-month LIBOR and the 3-month overnight indexed swap rate). While VIX reflects the volatility of the stock market, the TED spread and the LIBOR-OIS spread measure the credit risk of banks. In this section, we compare PNBO with these measures of funding constraints.

We first run OLS regressions of PNBO on the funding constraint measures. As Panel A of Table 8 shows, PNBO is significantly positively related to the TED spread and LIBOR-OIS spread. This positive relation is mainly due to the fact that PNBO rose significantly along with the TED spread (and the LIBOR-OIS spread) during the early part of the financial crisis. Subsequently, while PNBO turned significantly negative, the TED spread fell to and remained at low levels.[9]

In multivariate regressions, VIX, FG, and $Noise$ are all significantly negatively related to PNBO. In the quarterly regression, PNBO is significantly positively related to the growth rate in broker-dealer leverage Δlev, with an R^2 of 25%. These results suggest that when funding constraint tightens, the financial intermediaries tend to sell less of the DOTM SPX puts to public investors and might even become net buyers of these options.

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 the year-over-year change in broker-dealer leverage (Δlev) has strong predictive power for excess returns on stocks, corporate bonds, and treasuries. In a univariate regression (unreported), Δlev indeed predicts future market excess returns with a significant negative coefficient in our sample period. However, in a bivariate regression with PNBO, the coefficient on Δlev becomes insignificant, while that on PNBO remains significant. The R^2 of the bivariate regression is essentially identical to that in the univariate regression for PNBO. Similarly, when the other funding constraint measures are used in place of Δlev,

[9] The TED spread and LIBOR-OIS could become lower because of the cautionary measures banks take to reduce their risk exposures, which includes aggressively buying protections via DOTM puts and deleveraging, but they do not necessarily imply that banks are no longer constrained.

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

Panel A reports the results of the OLS regressions of PNBO 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; VIX is the CBOE VIX index; FG is the funding liquidity measure by Fontaine and Garcia (2012); $Noise$ is the illiquidity measure by Hu, Pan, and Wang (2013); Δlev is the balance sheet growth measure by Adrian, Moench, and Shin (2010). Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). The sample period is 1991 – 2012, except for the regressions with LIBOR-OIS, which is 2002 – 2012.

	A. Explaining PNBO with other measures of funding constraints						
TED	45.24					101.98	
	(13.58)					(17.18)	
LIBOR-OIS		0.44					
		(0.25)					
VIX			-0.27			-1.42	
			(0.88)			(0.70)	
FG				-4.89		-17.80	
				(3.90)		(3.87)	
$Noise$					0.55	-5.79	
					(3.30)	(2.62)	
Δlev							63.46
							(14.78)
Adj. R^2	0.10	0.05	0.00	0.00	0.00	0.26	0.25
Obs	264	133	264	264	264	264	88

	B. Predicting returns with PNBO and other measures of constraints						
PNBO	-61.16	-60.40	-65.12	-64.14	-64.98	-51.39	-72.22
	(13.89)	(13.65)	(16.42)	(16.43)	(14.69)	(14.19)	(21.45)
TED	-1.76					-4.14	
	(2.14)					(2.30)	
LIBOR-OIS		-0.05					
		(0.02)					
VIX			0.03			0.21	
			(0.10)			(0.11)	
FG				0.67		1.28	
				(0.70)		(0.71)	
$Noise$					-0.33	-0.48	
					(0.35)	(0.44)	
Δlev							-3.57
							(2.79)
Adj. R^2	0.17	0.31	0.17	0.17	0.18	0.20	0.17
Obs	264	133	264	264	264	264	88

the coefficient on PNBO is essentially unaffected.

We also further examine the relationship between PNBO and Δlev using Granger causality. We do this by forming a bivariate VAR and testing for the significance of either PNBO in predicting future changes in broker-dealer leverage or vice versa. In both cases, we find evidence that Granger causality runs both ways: PNBO incrementally predicts future changes in broker-dealer leverage and changes in leverage predict future changes in PNBO. We reject the null of no Granger causality (by a Wald test that all VAR coefficients are zero) at the 2.5% level or lower in all cases and this hold for either the VAR with fixed $p = 1$ lag or when we use the AIC or BIC criterion to optimally select the lag ($p = 7$ and $p = 4$, respectively.) We note also that there is little evidence that market returns Granger cause PNBO; the p-values in the associated hypothesis tests are all 0.68 or above for the different specifications.

2.7 Who sold the DOTM puts in the crisis?

As Figure 1 shows, the amount of DOTM SPX puts that public investors sold to the broker-dealers and market-makers in the period following the Lehman bankruptcy is quite large. To understand the risk sharing mechanism during times of distress, it is important to know who among the public investors sold the crash insurance to the constrained financial intermediaries. The SPX volume data from CBOE do not separate trades of retail investors from those of institutional investors. We use two strategies to answer this question. First, we compare the trading activities of the public investors in SPX options with those in SPY options. Second, we compare the trading activities of large against small orders in SPX options.

While SPX and SPY options have essentially identical underlying asset, it is well known among practitioners that SPX option volume has a significantly higher percentage of institutional investors. Compared to retail investors, institutional investors prefer SPX options more 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. Thus, as in SPX options, we construct

Table 9: **Comparing SPX vs. SPY Trades and Large vs. Small SPX Trades**

This table reports the results of the return forecasting regressions based on four measures of public net buying-to-open volume: PNBO for SPX options ($PNBO_{SPX}$), PNBO for SPY options ($PNBO_{SPY}$), PNBO for large orders on SPX options ($PNBO_{large}$), and PNBO for small orders on SPX options ($PNBO_{small}$). All options under consideration are DOTM, as defined by $K/S \leq 0.85$. r_{t+k} indicates market excess return in the kth month ahead. $r_{t\to t+k}$ indicates cumulative k-month market excess return. Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation based on Hansen and Hodrick (1980). Sample period for the SPX vs. SPY comparison: 2005/01 - 2012/12. Sample period for the large vs. small SPX trade comparison: 1991/01 - 2012/12.

Return	b_r	$\sigma(b_r)$	R^2	b_r	$\sigma(b_r)$	R^2
	$PNBO_{SPX}$			$PNBO_{SPY}$		
$r_{t\to t+1}$	-23.39	(6.26)	0.16	1.86	(0.74)	0.03
$r_{t+1\to t+2}$	-18.92	(5.17)	0.10	0.10	(1.35)	-0.01
$r_{t+2\to t+3}$	-24.42	(5.73)	0.17	0.05	(0.94)	-0.01
$r_{t\to t+3}$	-65.98	(16.00)	0.35	2.12	(0.97)	0.00
	$PNBO_{large}$			$PNBO_{small}$		
$r_{t\to t+1}$	-21.52	(8.41)	0.05	-12.94	(7.98)	0.00
$r_{t+1\to t+2}$	-16.21	(7.84)	0.03	-12.50	(8.39)	0.00
$r_{t+2\to t+3}$	-17.55	(9.41)	0.03	-16.11	(10.62)	0.01
$r_{t\to t+3}$	-54.53	(22.59)	0.10	-41.60	(26.56)	0.02

PNBO$_{SPY}$ for SPY options. Our SPY options volume data are from the CBOE and ISE, and cover the period from 2005/01 to 2012/12. Unlike the SPX options which trade exclusively on the CBOE, the SPY options are cross-listed at several option exchanges. Our SPY PNBO variable aggregates the volume data from the CBOE and ISE, which account for about half of the total trading volume for SPY options.

During the period of 2005/01 to 2012/01, PNBO$_{SPY}$ is positive in most months, suggesting that the public investors in the SPY market have been consistently buying DOTM puts. From 2008/09 to 2010/12, PNBO$_{SPX}$ (or PNBO) is negative in 22 out of 28 months, whereas PNBO$_{SPY}$ is negative in just 7 of the months. The correlation between PNBO$_{SPY}$ and PNBO$_{SPX}$ during this period is -0.27.

We also compare the ability of PNBO_{SPY} and PNBO_{SPX} in predicting market excess returns. As Table 9 shows, PNBO_{SPY} predicts one-month ahead market excess return positively in the period of January 2005 to December 2012. The coefficient on PNBO_{SPY} is also significantly positive when predicting three-month returns, but the R^2 is essentially zero. In the same period, the coefficient on PNBO_{SPX} is significantly negative at all four horizons, and the R^2 are much higher.

Next, we compare the predictive power of PNBO based on orders of different sizes. The SPX option volume data classify trades into large orders (more than 200 contracts per trade), medium orders (between 100 and 200 contracts), and small orders (less than 100 contracts). To the extent that institutional investors tend to execute large orders while retail investors trade in small orders, a comparison of PNBO_{large} and PNBO_{small} can also reveal the different behaviors of the two groups of public investors. Table 9 shows PNBO_{large} has significant predictive power for future market returns. While PNBO_{small} is also negatively related to future market returns, the relation is statistically insignificant, and the R^2 is much smaller. These comparisons suggest that it is the institutional investors who sold the DOTM put options to the financial intermediaries during the crisis period.

3 A Dynamic Model

In Section 2, we present empirical evidence connecting the trading activities in the market for DOTM SPX puts to option pricing, market risk premium, and various measures of intermediary constraints. In particular, there is time-varying equilibrium demand for crash insurance from public investors. The equilibrium demand is inversely related to the relative price of DOTM put options—times in which the equilibrium demand is low are generally times when the protection is very expensive. The demand for crash insurance is also informative about future stock market returns over and above the information in standard macro and financial variables. Finally, the demand for crash insurance is positively related to the changes in broker-dealer leverage. We now examine an equilibrium model consistent with these empirical facts.

3.1 Model setup

We consider an aggregate endowment in the economy which follows a jump diffusion process where the endowment is subject to both a diffusive risk and a jump risk. In particular, sudden severe drops in the aggregate endowment are a source of disaster risk in this economy. There are two types of agents in the economy: small public investors and competitive dealers. We assume there exists a representative public investor, who is denoted by agent P, and a representative dealer, denoted by agent D. To induce the two types of agents to trade, we assume that they have different beliefs about the probability of disasters. Such differences in beliefs capture in reduced form the advantages that dealers have in bearing disaster risk, whether it is due to differences in technology, agency problems, or behavioral biases.[10]

Specifically, we assume that both agents believe that the log aggregate endowment, $c_t = \log C_t$, follows the process

$$dc_t = \bar{g}dt + \sigma_c dW_t^c - \bar{d}\, dN_t, \tag{4}$$

where $\bar{g}$ and σ_c are the expected growth rate and volatility of consumption without jumps, W_t^c is a standard Brownian motion under both agents' beliefs, and $\bar{d}$ is the constant size of consumption drop in a diaster[11]. N_t is a counting process whose jumps arrive with stochastic intensity λ_t under the public investors' beliefs, and λ_t follows

$$d\lambda_t = \kappa(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda \sqrt{\lambda_t} dW_t^\lambda, \tag{5}$$

where $\bar{\lambda}$ is the long-run average jump intensity under P's beliefs, and W_t^λ is a standard Brownian motion independent of W_t^c. In general, the dealers are more willing to bear the disaster risk because (they act as if) they are more optimistic about disaster risk.

[10] Examples include government guarantees to large financial institutions and compensation schemes that encourages managers to take on tail risk. See e.g., Lo (2001), Malliaris and Yan (2010), Makarov and Plantin (2011).

[11] As in Chen, Joslin, and Tran (2012), one could generalize the model by allowing disaster size to have a time-invariant distribution.

We assume that they believe that the disaster intensity is given by $\rho\lambda_t$ with $\rho < 1$. We summarize the public investors' beliefs with the probability measure $\mathbb{P}_P$, and the dealers' beliefs with the probability measure $\mathbb{P}_D$.

Public investors have standard constant relative risk aversion (CRRA) utility:

$$U^P = E_0^P \left[\int_0^\infty e^{-\delta t} \frac{C_{P,t}^{1-\gamma}}{1-\gamma} dt \right], \tag{6}$$

where we focus on the cases where $\gamma > 1$. The superscript P reflects that the expectations are taken under the public investors' beliefs.

The utility function of the dealers are different. We assume that the dealers face an intermediation constraint that we model in a reduced form directly in terms of their utility. Specifically, we suppose that

$$U^D = E_0^D \left[\int_0^\infty e^{-\delta t} \frac{C_{D,t}^{1-\gamma}}{1-\gamma} e^{-\sum_{n=1}^{N_t} (\alpha_{\tau(n)} - \bar{\alpha})} dt \right], \tag{7}$$

where α_t is a stochastic variable representing the ability of the dealer to intermediate disaster risk. Limited ability to intermediate risk is modeled as increased risk aversion against market crashes. This specification generalizes the state-dependent preferences proposed by Bates (2008) in that it allows the dealers' risk aversion against crashes to rise with the probability of disasters.

Specifically, $\tau(n)$ is the time of the n$^{\text{th}}$ disaster since $t = 0$, $\tau(n) \equiv \inf\{s : N_s = n\}$ Thus, this crash-aversion term remains constant in between disasters. Suppose the dealer's log consumption drops by $d_{D,\tau(n)}$ at the time of the n$^{\text{th}}$ disaster. Then, at the same time, the marginal utility of the dealer jumps up by

$$e^{\gamma d_{D,\tau(n)} - (\alpha_{\tau(n)} - \bar{\alpha})} = e^{\left(\gamma - \frac{\alpha_{\tau(n)} - \bar{\alpha}}{d_{D,\tau(n)}}\right) d_{D,\tau(n)}}, \tag{8}$$

which implies that the dealer's effective relative risk aversion against the disaster is

$$\gamma_{D,\tau(n)} = \gamma - \frac{\alpha_{\tau(n)} - \bar{\alpha}}{d_{D,\tau(n)}}. \tag{9}$$

Thus, when $\alpha_t > \bar{\alpha}$, the dealer will have lower aversion to disaster risk than the public investor. As α_t falls, the dealer's effective risk aversion rises.

The intermediation capacity of the dealer may be related to the disaster intensity. We model the intermediation as being driven jointly by the disaster intensity, λ_t, and an independent factor, x_t, so that $\alpha_t = -a\lambda_t + bx_t$. Thus when $a > 0$, the intermediation capacity goes down as the intensity rises and the dealer becomes more averse to disaster risk.

Any jointly affine process for (c_t, λ_t, x_t) would be suitable for a tractable specification. For example, we could suppose that x_t follows an independent CIR process:

$$dx_t = \kappa_x(\bar{x} - x_t)dt + \sigma_x\sqrt{x_t}dW_t^x. \tag{10}$$

In our calibrations, we will choose the simple specification with $b = 0$ so that the intermediation capacity is perfectly correlated with the disaster intensity.

The main motivation for the dealer's time-varying aversion to crash risk is the time-varying constraint faced by financial intermediaries. Rising crash risk in the economy raises the intermediaries' capital/collateral requirements and tightens their constraints on tail risk exposures (e.g., Value-at-Risk constraints), which make them more reluctant to provide insurance against disaster risk. For example, see Adrian and Shin (2010) and He and Krishnamurhty (2012). In this sense, the shocks to the disaster intensity in the model also serve the purpose of generating time variation in the intermediation capacity of the dealer. We can further generalize the specification by making the dealer's aversion to crash risk driven by adding independent variations in the intermediation shocks.

We also assume that markets are complete and agents are endowed with some fixed share of aggregate consumption ($\theta_P, \theta_D = 1 - \theta_P$). The equilibrium allocations can be characterized as the solution of the following planner's problem, specified under the probability measure $\mathbb{P}_P$,

$$\max_{C_t^P, C_t^P} E_0^P \left[\int_0^\infty e^{-\delta t} \frac{(C_t^P)^{1-\gamma}}{1-\gamma} + \zeta \eta_t e^{-\delta t} \frac{(C_t^D)^{1-\gamma} e^{a \sum_{n=1}^{N_t} (\lambda_{\tau(n)} - \bar{\lambda})}}{1-\gamma} dt \right], \tag{11}$$

subject to the resource constraint $C_t^P + C_t^D = C_t$. Here, ζ is the the Pareto weight for the dealers and

$$\eta_t \equiv \frac{d\mathbb{P}_D}{d\mathbb{P}_P} = \rho^{N_t} e^{(1-\rho) \int_0^t \lambda_s ds}. \tag{12}$$

where $\rho = \bar{\lambda}_D / \lambda$, the relative likelihood of a jump under the two beliefs. From the first order condition and the resource constraint, we obtain the equilibrium consumption allocations $C_t^P = f^P(\tilde{\zeta}_t) C_t$ and $C_t^D = (1 - f^P(\tilde{\zeta}_t)) C_t$, where

$$\tilde{\zeta}_t = \rho_t^N e^{(1-\rho) \int_0^t \lambda_s ds + \alpha \sum_{n=1}^{N_t} (\lambda_{\tau(n)} - \bar{\lambda})} \zeta \tag{13}$$

and

$$f^P(\tilde{\zeta}) = \frac{1}{1 + \tilde{\zeta}^{\frac{1}{\gamma}}}. \tag{14}$$

The stochastic discount factor under P's beliefs, M_t^P, is given by

$$M_t^P = e^{-\rho t} (C_t^P)^{-\gamma} = e^{-\delta t} f^P(\tilde{\zeta}_t)^{-\gamma} C_t^{-\gamma}. \tag{15}$$

We can solve for the Pareto weight ζ through the lifetime budget constraint for one of the agents (Cox and Huang (1989)), which is linked to the initial allocation of endowments.

Our key focus will be on risk premiums and on the net public purchase of crash insurance which we relate to the market for deep out of the money puts in our empirical analysis. The risk premium for any security under each agent's beliefs is the difference between the expected return under $\mathbb{P}_i$ and under the risk-neutral measure $\mathbb{Q}$.

$$E_t^i[R^e] = \gamma \sigma_c \partial_B P + (\lambda_t^i - \lambda_t^{\mathbb{Q}}) E_t^d[R], \quad i = D, P, \tag{16}$$

where we use the shorthand that $\partial_B P$ denotes the sensitivity of the security to Brownian shocks and $E_t^d[R]$ is the expected return of the security *conditional on a disaster*. Since consumption will be relatively smooth in our calibration, the return of securities which are not highly levered on the Brownian risk will be dominated by the jump risk term. Moreover, agents agree about the Brownian risk and have the same risk aversion with respect to these shocks so there will be no variation in the Sharpe ratio for Brownian risk. In light of these facts, we focus on the variation in the jump risk premium, as measured by $\lambda^{\mathbb{Q}}/\lambda_P^{\mathbb{P}}$.

The stochastic discount factor characterizes the unique risk neutral probability measure $\mathbb{Q}$ (see, e.g., Duffie 2001). The risk-neutral disaster intensity, $\lambda_t^{\mathbb{Q}} \equiv E_t^d[M_t^i]/M_t^i \lambda_t^i$, is determined by the expected jump size of the stochastic discount factor at the time of a disaster. When the risk-free rate and disaster intensity are close to zero, the risk-neutral disaster intensity has the nice interpretation of (approximately) the value of a one-year crash insurance contract that pays one at $t+1$ when a disaster occurs between t and $t+1$. In our setting, the risk-neutral jump intensity is given by

$$\lambda_t^{\mathbb{Q}} = e^{\gamma \bar{d}} \frac{\left(1 + (\rho \tilde{\zeta}_t)^{\frac{1}{\gamma}}\right)^{\gamma}}{(1 + \tilde{\zeta}_t^{\frac{1}{\gamma}})^{\gamma}} \lambda_t. \tag{17}$$

In order to define the market size, we must consider how the Pareto efficient allocation is obtained. The equilibrium allocations can be implemented through competitive trading in a sequential-trade economy. Extending the analysis of Bates (2008), we can consider four types of traded securities: (i) a risk-free money market account, (ii) a claim to aggregate consumption, and (iii) a crash insurance contracts which pays one dollar in the event of a disaster in exchange for a continuous premium, and (iv) a separate instrument sensitive only to shocks in the disaster intensity. As in Chen, Joslin, and Tran (2012), since agents agree about the Brownian risk and have identical aversion to the risk, they will proportionally hold the Brownian risk according to their consumption share. With the instruments we have specified, this means they will proportionally hold the consumption claim. Thus, the agents will hold proportional exposure to the disaster risk from their exposure to the consumption claim. Motivated by these facts, we define the net public

Table 10: **Model Parameters**

risk aversion: γ	4
time preference: δ	0.03
mean growth of endowment: $\bar{g}$	0.025
volatility of endowment growth: σ_c	0.02
mean intensity of disaster: $\bar{\lambda}$	1.7%
speed of mean reversion for disaster intensity: κ	0.142
disaster intensity volatility parameter: σ	0.05
dealer risk aversion parameter: α	1.0

purchase for crash insurance as the (scaled) difference between the consumption loss the public bears in equilibrium minus the consumption loss that the public would bear without insurance. That is, the public purchase for insurance is the difference between $e^{-\bar{d}}(f^P(\tilde{\zeta}_t^d) - f^P(\tilde{\zeta}_{t-}))$ (where $\tilde{\zeta}_t^d$ is the value of $\tilde{\zeta}_t$ conditional on a disaster occurring at time t: $\tilde{\zeta}_t^d = \rho e^{\alpha(\lambda_{t-}-\bar{\lambda})}\tilde{\zeta}_{t-}$) and $e^{-\bar{d}}(f^P(\tilde{\zeta}_t^d) - f^P(\tilde{\zeta}_{t-})$. Thus we define the net public purchase for insurance as

$$\text{net public purchase for crash insurance} = e^{-\bar{d}}\big(f^P(\tilde{\zeta}_t \rho e^{\alpha(\lambda_{t-}-\bar{\lambda})}) - f^P(\tilde{\zeta}_t)\big). \tag{18}$$

3.2 Net public purchase and risk premia in the dynamic model

We now study the relationship between public purchase and risk premia in the context of our dynamic model. We calibrate our model as in Chen, Joslin, and Tran (2012) and Wachter (2012). The key new parameter that we introduce is the time-variation in aversion to jump risk. We parameterize this by setting $a = \alpha\bar{d}/\sigma_{SS}(\lambda)$, where $\sigma_{SS}(\lambda)$ is the volatility of the stationary distribution of λ. We choose $\alpha = 1$, which together with the other parameters implies that when $\lambda = 2.35\%$ (one standard deviation from the steady state volatility (65 bp) above the long run mean (1.7%)), an economy populated by only the dealer will behave as an economy with a representative agent with relative risk aversion of 5 with respect to jumps (one higher than if he had standard CRRA utility with $\gamma = 4$). As a baseline comparison, we also consider the parameterization with $\alpha = 0$, which corresponds to the stochastic intensity model of Chen, Joslin, and Tran (2012).

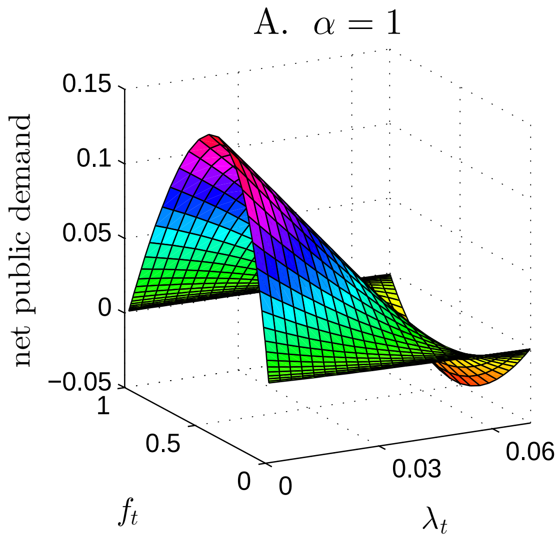


Figure 6: **Net Public Purchase for Crash Insurance.** The two panels plot the net public purchase for crash insurance as a function of the public investor P's consumption share (f_t) and the disaster intensity under P's beliefs (λ_t). Panel A considers the case when $\alpha = 1$, which implies that a 1-standard deviation increase in the disaster intensity from its long-run mean effectively increases the dealers' relative risk aversion against disasters by 1 in the homogeneous-agent economy. Panel B considers the case when $\alpha = 0$.

In our model, there are two state variable: λ_t, the likelihood of a disaster, and f_t, the public investors consumption share. Figure 6 plots the net public purchase of crash insurance as a function of the public consumption share and the jump intensity. When $\alpha = 0$ (panel B), the amount of risk sharing does not depend on λ as the motivation to share risk depends only on the size of the jump in consumption. The equilibrium public purchase is close to zero when either the public has a low consumption share (the public has limited resources to buy insurance) or when the public has high consumption share (the dealers have limited ability to provide insurance) with a peak in the middle where the public and dealers share a lot of risk. In contrasts, Panel B shows that when the dealers have time-varying aversion to jump risk, the relationship is much more complex. For low levels of the intensity the patterns is the same as before since the dealers are as willing (or even more willing) to sell crash insurance. However, as λ rises, the dealer becomes more averse to jump risk and is less willing to provide insurance. When the intensity becomes high enough, the dealers become so averse to jump risk that they even begin to become

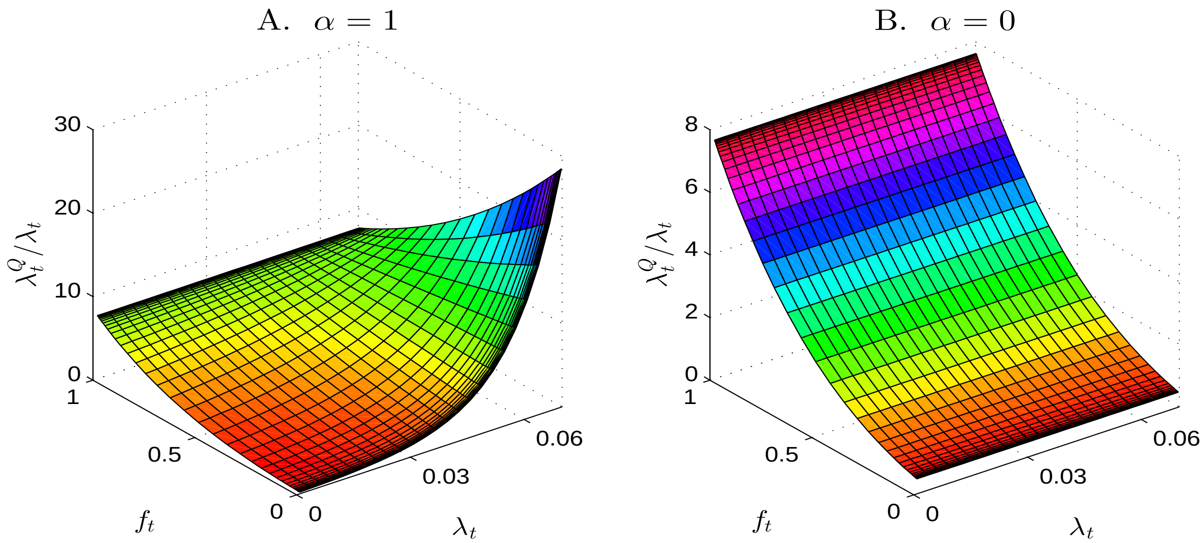


Figure 7: **Disaster Risk Premium.** The two panels plot the conditional disaster risk premium as a function of the public investor P's consumption share (f_t) and the disaster intensity under P's beliefs (λ_t). Panel A considers the case when $\alpha = 1$. Panel B considers the case when $\alpha = 0$. Again, $\alpha = 1$ means that a 1-standard deviation increase in the disaster intensity from its long-run mean effectively increases the dealers' relative risk aversion against disasters by 1 in the homogeneous-agent economy.

buyers of insurance rather than sellers.

Figure 7 plots the jump risk premium, as measured by $\lambda^{\mathbb{Q}}/\lambda$, as a function of the public consumption share and the jump intensity. In the case of $\alpha = 0$, the jump risk premium rises as there are fewer dealers to hold the jump risk. When $\alpha = 1$ and λ is low, the jump risk premium falls as the dealer's consumption share increases. When λ is high enough, this relationship reverses and the premium rises as the dealer gains consumption share. The reason for this is that as λ rises, the dealer becomes more risk averse and eventually demands a higher premium than the public; this relation can be seen be following the curve with $f_t = 0$, corresponding to the case where there is only the dealer.

Next, we examine the relation between the net public purchase of crash insurance and the disaster risk premium in equilibrium. We do so by first holding constant the consumption share of the public investors (f_t) while letting the disaster intensity (λ_t) vary over time. The results are in Figure 8. When $\alpha = 1$, for each of the consumption shares considered ($f_t = 0.9, 0.8, 0.5$), the model predicts a negative relation between the net

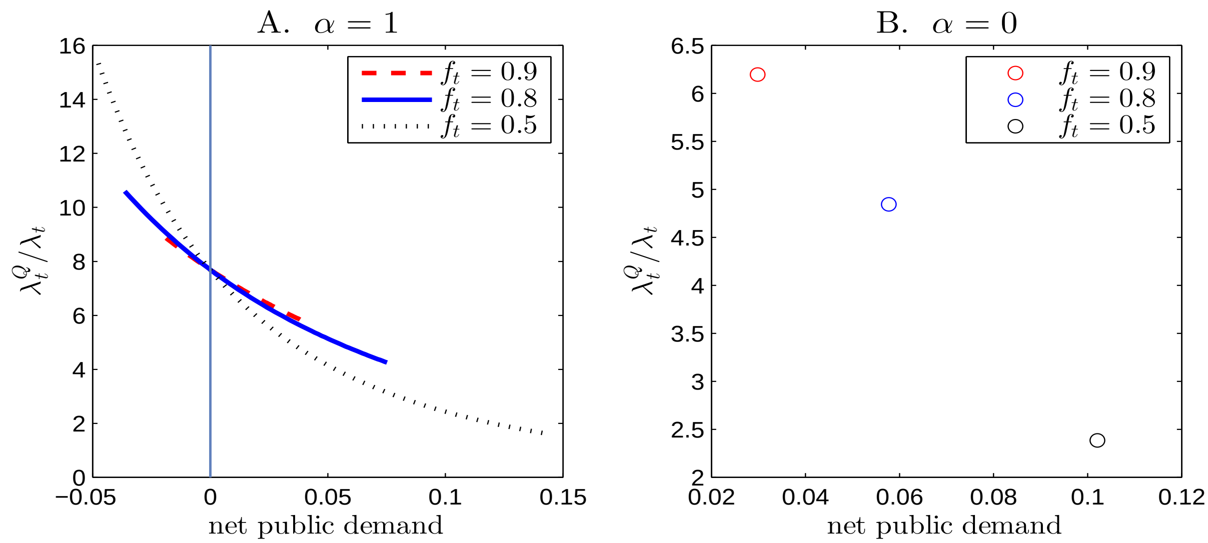


Figure 8: **Net Public Purchase for Crash Insurance and Risk Premium: Fixed Consumption Share.** The two panels plot the conditional disaster risk premium against the net public purchase for crash insurance while holding the public investors' consumption share f_t constant. Panel A considers the case when $\alpha = 1$. Panel B considers the case when $\alpha = 0$.

public purchase for crash insurance and risk premium. This negative relation is consistent with our empirical finding of $b_{VP} < 0$ in Equation (1) and $b_r < 0$ in Equation (3).

In contrast, when $\alpha = 0$, i.e., when the dealer has constant risk aversion for disaster risk, both the net public purchase and the disaster risk premium remain constant as the disaster intensity changes. This comparison again highlights the key role played by the dealer's time-varying aversion to disaster risk in our model.

When we fix the disaster intensity and let the consumption share vary over time, the relation between the net public purchase of crash insurance and the disaster risk premium is no longer monotonic. Consider first the case with $\alpha = 0$, i.e., the case where the dealer has constant relative risk aversion (see Panel B of Figure 9). In this case, regardless of the disaster intensity, there is a unique relation between the two quantities: as consumption share of the public investors changes from 0 to 1, the net public purchase, as defined in (18), starts at 0, reaches its peak at 11%, and then falls back to 0 eventually. The limiting case where the public investor own all the aggregate endowment is marked by the red circle on the y-axis. In this process, because the relative amount of risk sharing by the

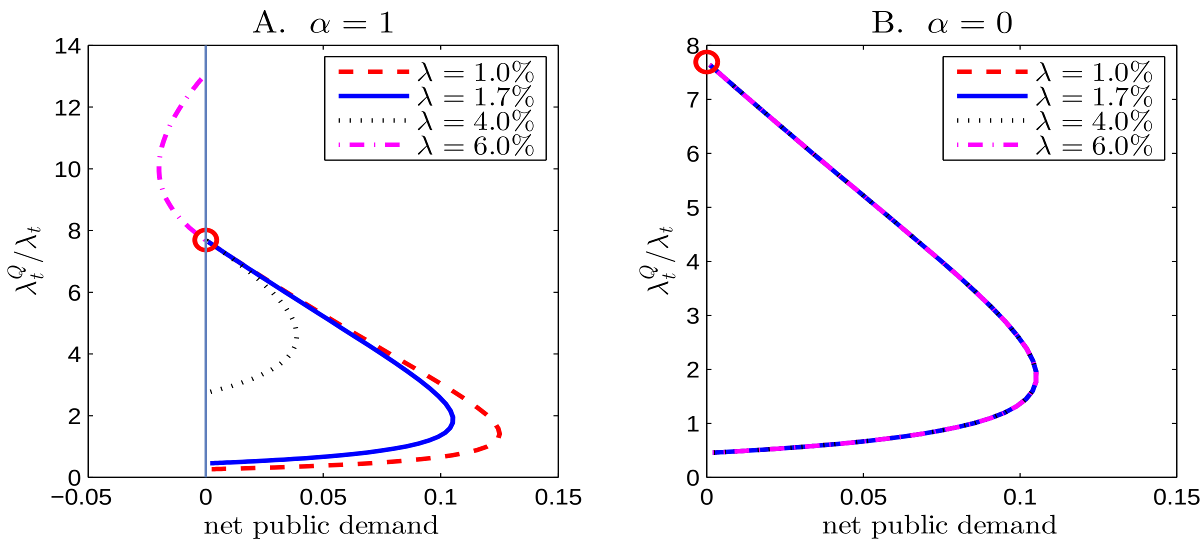


Figure 9: **Net Public Purchase for Crash Insurance and Risk Premium: Fixed Disaster Intensity.** The two panels plot the conditional disaster risk premium against the net public purchase for crash insurance while holding the disaster intensity λ_t constant. Panel A considers the case when $\alpha = 1$. Panel B considers the case when $\alpha = 0$. The red circles mark the limiting cases in the two economies where public investors own all the aggregate endowment.

public investor falls as he gains a larger share of consumption, the disaster risk premium in equilibrium rises monotonically until it reaches the limit of $e^{\gamma \bar{d}} = 7.7$.

Consider now the case where the dealer has time-varying risk aversion ($\alpha = 1$). When $\alpha = 1$, (Panel A of Figure 9) the relation between the net public purchase of crash insurance and the disaster risk premium is qualitatively similar to that in Panel B (where $\alpha = 0$) when the disaster intensity is not too high. The equilibrium where public investors own all the endowment is still identical regardless of the level of disaster intensity (again marked by the red circle), but the other extreme where the dealer owns all the endowment has different risk premiums for different disaster intensities. This result is simply due to the dealer's time-varying risk aversion. When λ_t is sufficiently high, the dealer can become so averse to disaster risk that, despite his optimist beliefs about the chances of disasters, the dealer still demands a higher premium than the public investors would. For this reason, when λ_t is sufficiently high, the net public purchase of crash insurance turns negative, i.e., the public investor is now insuring the dealer against disasters, and the disaster risk

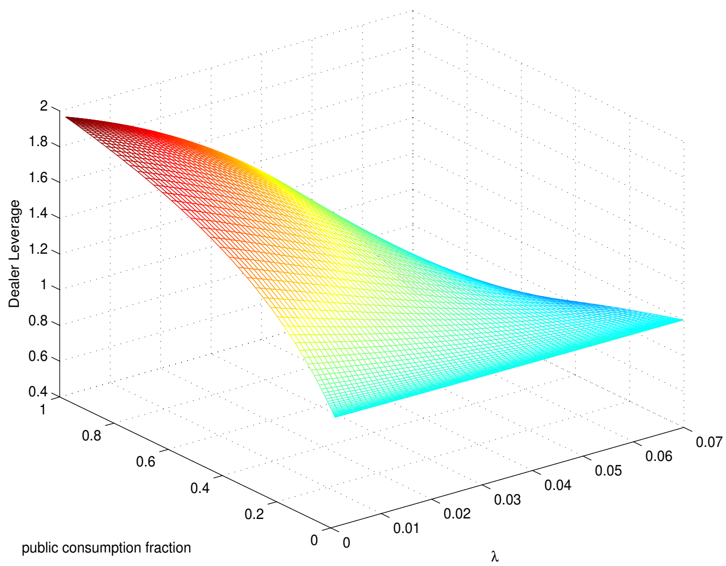


Figure 10: **Dealer leverage.** This figure plots the dealer leverage as a function of the public consumption fraction and the disaster probability.

premium in the economy exceeds the highest level in the case with constant risk aversion.

Panel A of Figure 9 also illustrate that, as the level of disaster intensity λ_t rises, the relation between the disaster risk premium and the net public purchase of crash insurance becomes more negative in the region near the red circle. That is, when public investors own the majority of the wealth in the economy, changes in the public demand for crash insurance are accompanied by larger swings in aggregate risk premium when the level of crash risk is higher. This prediction is confirmed by our empirical finding of $c_{VP} < 0$ in Equation (1).

Figure 10 examines the relationship between the dealer leverage and the consumption share of the public investor and the disaster intensity. Here, we measure the leverage of the dealers as the ratio of the equilibrium consumption loss of the dealers (which includes losses from selling disaster insurance and their equity position) to their consumption loss they would bear without selling any insurance. When the disaster probability is low, so that the dealers are relatively unconstrained and aggressively sell insurance, the dealer leverage increases to a maximum of around two when public investors own the majority of the wealth. For a given level of disaster probability, the dealer leverage decreases as public investors' wealth (consumption) share becomes smaller. As the disaster probability rises, the dealers become more constrained and begin to deleverage. In some cases, the dealer leverage as we define may even fall below one, indicating that the dealers are acquiring insurance to reduce their exposures to disaster risks.

Taken together, the results from Figure 6 and Figure 10 show that our model is able to capture the negative relation between intermediary leverage growth and the net quantity of trading in disaster insurance in the data, as documented in Table 8.

3.3 Extension

Our main model presented in Section 3.1 captures a number of the key features we have found in the data. In particular, the model captures the fact that when equilibrium public buying is low, risk premia may be high as this may correspond to time when dealers are (or act as if they are) more risk averse. However, as in Chen, Joslin, and Tran (2012), wealth moves slowly between the public sector and dealers outside of disasters and only through crash insurance premiums. In this section, we generalize our main model to account for more general time variation in the relative wealth of the public and dealers.

Consider the case where the public and dealer not only view the disaster events differently, but also disagree about the future path of the likelihood of disasters. Specifically, consider the more general form of (12) where

$$\frac{d\mathbb{P}_D}{d\mathbb{P}_P} = \rho^{N_t} e^{(1-\rho)\int_0^t \lambda_s ds} \times e^{-\int_0^s \theta_s dW_s^{\lambda} - \int_0^t \theta_s^2 ds}. \tag{19}$$

and θ_s is some process satisfying Novikov's condition. For example, with an appropriate chose of θ_t, we may have that the dealer will believe that the dynamics of the λ_t are

$$d\lambda_t = \kappa^D(\bar{\lambda}^D - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t^{\lambda,D},$$

where $W_t^{\lambda,D}$ is a standard Brownian motion under the dealer's beliefs.

An example we have in mind is that the dealer may believe that when the intensity is

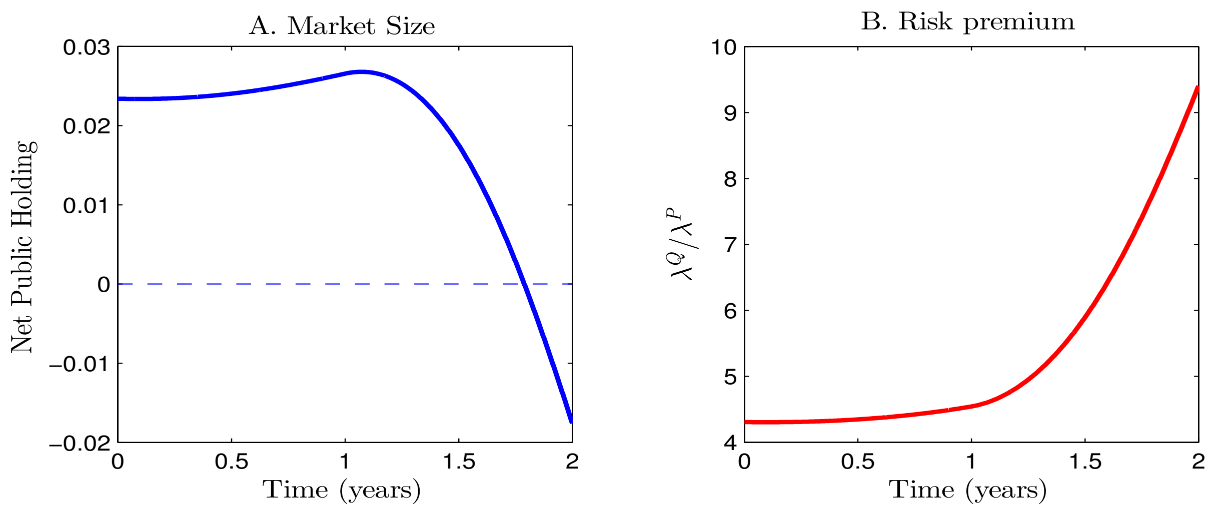


Figure 11: **Dealer constraint and derivative supply.** This figure plots the equilibrium holding of crash insurance by the public investors (Panel A) and disaster risk premium (Panel B) for a hypothetical history where the intensity rises from 1.7% to 3.7% over a two year period. The public has an initial consumption fraction of 0.25. The dealer's relative risk aversion is initially $\gamma = 4$ and then rises in the second half quadratically to $\gamma = 6.5$.

high, it will mean revert more quickly to the steady state than it actually will. When this is the case, the dealer will make bets with the public that the intensity will fall. This will cause the public's relative wealth to grow if the intensity continues to rise. Thus even if the dealers are becoming more risk averse as the intensity rises, there will be a greater demand for crash protection from the public and in equilibrium the net effect can be that the size of the insurance market increases. Without this additional trading incentive, the relative wealth of the public and dealers will be nearly constant over short horizons. This extension allows us to capture some patterns seen in the crisis. In Figure 1, we saw that in the early stages of the crisis, the demand for crash insurance spiked and subsequently bottomed out as we reached the later stages.

The extended model can capture these types of features. To see this, we extend our base model with the additional assumption that the dealers believe that λ_t mean reverts ten times faster than the public (a half life of 0.48 years versus 4.8 years.) For simplicity, we assume that over a two year period the disaster intensity rises from its steady value of 1.7% at a rate of 1%/year to 3.7%. We initialize the public with a planner weight such that they initially represent 25% of consumption. We also model the implied risk aversion of the dealer to remain constant at $\gamma = 4$ in the beginning of the sample and then increase quadratically to $\gamma = 6.5$ at the end of the period. Figure 11 plots the resulting market size (Panel A) and risk premium (Panel B), as measured by $\lambda^{\mathbb{Q}} - \lambda^{\mathbb{P}}$. Generally, the patterns we see are qualitatively similar to those found in Figure 1. The public begins buying more insurance as the dealers lose money on their λ bets. As the crisis deepens, the dealers start to become very risk averse and the market dries up to the point where the dealers even become buyers of protection. Across this time period, the risk premium at first increases very slowly until the dealers are no longer willing to hold the risk and the premium begins to increase rapidly.

4 Conclusion

We provide evidence that the trading activities of financial intermediaries in the market of DOTM SPX put options are informative about the tightness of intermediary constraints. We use the price-quantity relation to isolate shocks to intermediaries' capacity to provide crash insurance. 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 price-quantity relation for DOTM puts is negative, 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 are consistent with the prediction of time-varying intermediary constraints driving the aggregate risk premium.

To explain these findings, we build a general equilibrium model of the crash insurance market. The model captures the time-varying intermediary constraints in reduced form, which provides the analytical tractability that makes it easy to examine the dynamic relations between the endogenous demand for crash insurance by public investors, the disaster risk premium, and intermediary leverage.

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, 2012, Financial intermediaries and the cross-section of asset returns, Working Paper.

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.

Adrian, Tobias, and Hyun Song Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418–437.

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.

Bakshi, Gurdip, and George Panayotov, 2013, Currency carry trade return predictability and asset pricing implications, *Journal of Financial Economics* pp. 139–163.

Bates, David S., 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,* forthcoming.

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, 2013, A macroeconomic model with a financial sector, Working Paper, Princeton University.

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

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.

Cox, John C., and Chifu Huang, 1989, Optimum consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 49, 33–83.

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, Lars Peter, and Robert J Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–53.

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

He, Zhiguo, and Arvind Krishnamurhty, 2012, Intermediary asset pricing, Forthcoming, American Economic Review.

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

Kelly, Bryan, 2012, Tail risk and asset prices, Working Paper, University of Chicago Booth School of Business.

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.

Lo, Andrew, 2001, Risk management for hedge funds: Introduction and overview, *Financial Analysts Journal* 57, 16–33.

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

Makarov, Igor, and Guillaume Plantin, 2011, Rewarding trading skills without inducing gambling, Working Paper, London Business School.

Malliaris, Steven, and Hongjun Yan, 2010, Reputation Concerns and Slow-Moving Capital, Working Paper, Yale University.

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

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